The international journal of science / 23 February 2023

nature



BRAVE NEW WORLDS

JWST offers unprecedented insights into exoplanets

Fighting back
Ukraine's scientists defy the odds to keep doing research

Smart decision
Why the world must ban autonomous lethal weapons

Climate crisis
Tree rings suggest drought helped end the Hittite empire

VISIT…

LANZAROTE Caliente.com

Nature.2023.02.25

[Sat, 25 Feb 2023]

- **This Week**
- **News in Focus**
- **Books & Arts**
- **Opinion**
- **Work**
- **Research**
- **Amendments & Corrections**

This Week

- **Rebuilding Ukrainian science can’t wait — here’s how to start** [22 Feb 2023]
  • International support has rightly focused on researchers who have been forced to flee. Ukraine’s devastated research infrastructure needs assistance, too.
- **Nature welcomes Registered Reports** [22 Feb 2023]
  • From this week, Nature will be publishing an additional type of research paper — designed to encourage rigour and replication.
- **Science needs strong advocates in the Nigerian elections** [21 Feb 2023]
  • Presidential candidates pay lip service to the role of science and technology in solving Nigeria’s problems — researchers must come together to be heard.

- EDITORIAL
- 22 February 2023

Rebuilding Ukrainian science can't wait — here's how to start

International support has rightly focused on researchers who have been forced to flee. Ukraine's devastated research infrastructure needs assistance, too.



The library at Taras Shevchenko National University of Kyiv was damaged last October by Russian missiles.Credit: Yevhen Kotenko/Ukrinform/NurPhoto/Shutterstock

The year since the Russian invasion of Ukraine began has been truly desperate for Ukrainian scientists. Many researchers left their laboratories to

join the front line. Some have made the ultimate sacrifice. Around 6,000 — one-tenth of all researchers in the country — have left. Some 80% of these are women, and the majority are now in Poland and Germany, with many more in France, Spain and Italy, and others in at least 16 further countries. A once fully functioning research system has been severely damaged, as *Nature* reports.

One year on from Russia's invasion, we're urging the international research community to prioritize support for not only individual researchers and their teams, but also Ukraine's science system as a whole. With the war far from over, Ukraine's diaspora of researchers must continue to be supported, but so must those who are holding the fort at home.

Some 27% (91) of Ukraine's 334 research and higher education institutes have suffered war damage; 4 institutes have been completely destroyed and 228 are so far unharmed. Both the renowned Kharkiv Institute of Physics and Technology and the world's largest decametre-wavelength radio telescope, at the Institute of Radio Astronomy in Kharkiv, were heavily damaged. Although many men aged 18–60 are conscripted, scientific research is considered an essential wartime occupation. Therefore many researchers, men and women, remain in post, often working remotely in safer parts of the country.



The fight to keep Ukrainian science alive through a year of war

International institutions and research networks have started to expand the scope of their support to include research in Ukraine itself. On 2 February, the European Commission announced that it would open a research office in Kyiv this year. This follows last June's decision to associate Ukraine to the EU's flagship funding scheme, Horizon Europe. As of January, Ukraine's researchers were already participating in 49 Horizon Europe projects, with a total funding of €13 million (US$ 13.9 million). But an association agreement gives Ukraine's researchers the same status as those from EU states, including the ability to lead projects.

Elsewhere, the Polish Academy of Sciences and the US National Academy of Sciences have jointly launched an innovative research-funding scheme for Ukrainian principal investigators and other researchers who are currently in Poland, but whose research teams can be based in Ukraine or elsewhere. Last year, Universities UK, which represents higher-education institutions in the United Kingdom, launched a scheme funded by the UK government that connects Ukrainian and UK universities, with an explicit aim that Ukrainian science must benefit. The scheme, which is part of a £5 million (US$6 million) investment, will support a range of activities, including remote research collaborations and the use of lab space for scientists who need to come to the United Kingdom temporarily to complete parts of their work that are currently impossible to finish in Ukraine. And last month, mathematicians from Ukraine, together with European maths research institutes, announced plans to create an International Centre for Mathematics in Ukraine. These are pioneering schemes that have Ukrainian science's long-term needs as their guiding principle.

Management science

But a more coordinated approach is needed, and one that supports not only the activity of science but also its organization and management. Last June, representatives of an international group of science academies, including ones in Ukraine, Europe and the United States, met in Warsaw and put together what they called action steps for rebuilding Ukraine's science, research and innovation. They rightly said that planning for a post-war science recovery of Ukraine should begin now. The group has created a forum, which it is calling the Ukrainian Science, Innovation, and Research

Coordinating Group, to share knowledge of its support for Ukraine's research system.

One priority for Ukraine and its international partners must be to review the country's system for organizing and funding research. Ukraine has a long-established science academy and the country's research strengths include agricultural technology, energy and planetary science.



[Seven ways the war in Ukraine is changing global science]

But the academy (rather than universities) is responsible for organizing and funding most research institutes, which is partly a legacy of how science was organized before Ukrainian independence in 1991, when the country was part of the Soviet Union. Research institutions were part of the state and lacked autonomy to make their own decisions, including which projects to fund, without the say-so of government officials. In the years before the war, Ukraine enacted some reforms such as the creation of an independent grant-giving body, the National Research Foundation of Ukraine, and an advisory body, the National Council of Ukraine on Science and Technology Development. But these organizations struggle to fulfil their roles.

Last November, the Ministry of Education and Science of Ukraine published a draft recovery plan for education and science. The plan includes a proposal to create a high-risk, high-reward funding agency along the lines of the US

Defense Advanced Research Projects Agency (DARPA). It also calls for closer integration with the European and wider global research communities, and to harness science in the task of post-war reconstruction — similar to the US-funded Marshall Plan, in which research and researchers were integral to the rebuilding of Europe after the Second World War.

There should always be a place for emergency help for Ukrainian scientists who have had to leave the country. But it's also time for deeper efforts to go towards displays of solidarity with researchers who remain in the country. The international science community should start planning how best to prepare the country's research infrastructure for the end of the war. Long-term partnerships that focus on capacity-building will be crucial, particularly in the areas of management, monitoring and policy. These collaborations must try to sustain day-to-day research as much as possible now, so that the research community can hit the ground running and be much more effective as soon as the conflict ends.

Nature **614**, 593-594 (2023)

doi: https://doi.org/10.1038/d41586-023-00505-3

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00505-3

| Section menu | Main menu |

- EDITORIAL
- 22 February 2023

Nature welcomes Registered Reports

From this week, *Nature* will be publishing an additional type of research paper — designed to encourage rigour and replication.



A Registered Report is a scientific paper that a journal commits to publishing if the research question and the methodology chosen to address it pass peer review.Credit: Getty

This year marks the 50th anniversary of *Nature*’s decision to mandate peer review for all papers. It’s an appropriate time to introduce readers and authors to Registered Reports, a research-article format that *Nature* is

offering from this week for studies designed to test whether a hypothesis is supported (see go.nature.com/3kivjh1).

The fundamental principle underpinning a Registered Report is that a journal commits to publishing a paper if the research question and the methodology chosen to address it pass peer review, with the result itself taking a back seat. For now, *Nature* is offering Registered Reports in the field of cognitive neuroscience and in the behavioural and social sciences. In the future, we plan to extend this to other fields, as well as to other types of study, such as more exploratory research.

Why are we introducing this format? In part to try to address publication bias, the tendency of the research system — editors, reviewers and authors — to favour the publication of positive over negative results. Registered Reports help to incentivize research regardless of the result. An elegant and robust study should be appreciated as much for its methodology as for its results.

Submitting a study

As for how it works, authors of Registered Reports are asked to make a pre-submission enquiry with their research plan before they embark on a study. Typically, this plan should include the research question being asked, an explanation of why this work fits the Registered Reports format and a brief explanation of the methods to be used and how data will be collected. The research question must meet *Nature*'s existing editorial criteria for scientific impact and the strength of the underpinning evidence. If these criteria are satisfied, the plan will be sent for peer review. Reviewers will judge submissions on the basis of a question's importance, either to a research field or more broadly (to an economy, the environment or society, for example). They will also assess the robustness of a study's design and analysis. If the reviewers are satisfied, the journal will commit to publishing the findings, as long as the methodology does not change during the course of a study.

To be clear, Registered Reports are not new. They have been around for at least a decade[1]. The format is already offered by a number of Nature

Portfolio journals, including *Nature Human Behaviour*, *Nature Methods, Nature Communications* and *Scientific Reports*. According to the Center for Open Science, more than 300 journals already offer this format, up from around 200 in 2019. But despite having been around for a while, Registered Reports are still not widely known — or widely understood — among researchers. This must change. And, at *Nature*, we want to play a part in changing it.



Good research begins long before papers get written

The lack of awareness is partly the result of systemic factors that favour the conventional research paper. Studies in which the main emphasis is on results are valuable — and are used in assessing promotions and grant applications. They are also important to institutions, which can use them to obtain funding, for example.

Clearly, more needs to be done to emphasize the benefits of Registered Reports. The format helps to reinforce the necessity of rigour in study design and methodology. At the same time, it provides in-built opportunities for feedback. Both of these help researchers to spot problems in studies before it is too late to fix them. Registered Reports are becoming a marker of quality — peers, institutions and funders are beginning to realize that the format represents a high standard of work[2]. Moreover, the format can make the peer-review process more constructive and amicable.

Collaborative potential

Registered Reports (along with other formats) might also have a role in helping to resolve disagreements within disciplines. For example, psychology researchers have disagreed on whether individuals' subjective experiences of emotion are influenced by their own facial expressions. Researchers with different views saw the Registered Reports format as a way to collaborate to test the hypothesis. Their findings[3] were published last October.

A decade ago, *Nature* developed a reporting summary for authors of manuscripts — a checklist in which authors are asked to state, for example, whether experimental findings have been replicated or whether a sample size is appropriate. Registered Reports are a progression towards greater emphasis on rigour and study design. They are also a format that recognizes both how science is done and that good research starts well before the paper is written.

Nature **614**, 594 (2023)

doi: https://doi.org/10.1038/d41586-023-00506-2

References

1. Chambers, C. D. & Tzavella, L. *Nature Hum. Behav.* **6**, 29–42 (2022).

2. Soderberg, C. K. *et al. Nature Hum. Behav.* **5**, 990–997 (2021).

3. Coles, N. A. *et al. Nature Hum. Behav.* **6**, 1731–1742 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00506-2

- WORLD VIEW
- 21 February 2023

Science needs strong advocates in the Nigerian elections



Presidential candidates pay lip service to the role of science and technology in solving Nigeria's problems — researchers must come together to be heard.

- Amina Ahmed El-Imam

Here in Nigeria, there have been interminable queues at fuelling stations nationwide for almost a year. I feel like a victim of robbery every time I buy petrol from the black market at twice the official rate. As a microbiology professor and working mother of four, I cannot navigate the chaos and long queues at the handful of stations that sell petrol at the government-approved price. Seared into my mind is the image of a brilliant engineer and a senior professor of medicine I know — whose time could be better spent doing

research to improve people's lives — queuing for hours outside a petrol station early on a Monday morning.

Usually, in the months ahead of general elections in Nigeria, elected leaders put on a semblance of governance to win votes. But now, even with the presidential and National Assembly elections this week, life in Africa's most populous country is arguably harder than in any other peacetime period. Science could be part of the solution — yet seemingly has no advocate in these elections.

Chief among our woes is the current cash shortage, which came on the heels of record-high inflation of more than 20% in August and September last year. The Central Bank of Nigeria mandated a rushed swap of high-value naira banknotes with insufficient new notes, ostensibly to stymie illegal activities. This has caused chaos at banks and fights in ATM queues — a man even died of unknown causes while queuing to get his bank card. Many Nigerians, 40% of whom have no bank account, have no choice but to keep struggling to obtain scarce replacement banknotes.



Could Africa be the future for genomics research?

On top of this, fuel shortages and perennial electricity outages are coupled with deficient technology infrastructure. Buggy, unsafe banking applications mean that many people have had to return to the archaic process of queuing in banks for the most mundane financial transactions.

Science can't immediately solve all our problems. But Nigerian politicians should heed the profound impact of investment in science and technology in countries such as South Korea and Malaysia. Our leaders need to break free of the trap of economic dependence on the export of commodities such as oil, minimize the activities of kleptomaniac bureaucrats and attract more investments to the science and technology sector. One of Nigeria's biggest problems is what social scientists call the 'resource curse', an observation supported by many studies over the years (X. Sala-i-Martin and A. Subramanian *J. Afr. Econ.* **22**, 570–615; 2013). In short, the discovery of oil has fuelled internal corruption and waste while stifling innovation.

But we do have home-grown advances in science and technology that can ameliorate some of these challenges. For instance, innovative Nigerians have invented solar-powered electric car,s and cookers that directly combust liquified natural gas that would help with the current petrol shortages and relieve the high cost of cooking fuel.

But the Federal Ministry of Science, Technology and Innovation has been unable to support commercialization of most of these inventions. This is not a surprise: the ministry's allocation in the 2023 federal budget is about 0.85%, close to what it has been for much of the past decade.



Female scientists in Africa are changing the face of their continent

One consequence of this stagnation is that we are losing our best young scientists. I have grown tired of watching some of my best students emigrate to other countries for PhD programmes and job opportunities, frustrated by the absence of research facilities and flat-lining incomes in a time of hyperinflation. In 2009, the government agreed to improve salaries for researchers – but more than a decade on, it has still failed to follow through.

Yet there is a worrying absence of conversation around science in the current election campaigns. Two of the three front-runners for the presidency mention science and technology in their manifestos, but have no specific plans for boosting scientific research and innovation, either in the manifestos or on the campaign trail. This seems to be a mere box-ticking exercise. Politicians are not afraid to make big-ticket promises because the electorate rarely demands accountability, so the near-total silence on science, technology and innovation is a worrying harbinger of future neglect.

We must collaborate to pull science in Nigeria out of the cold. The Nigerian Academy of Science, because of its influence with the federal government, is best situated to draw the candidates' attention to the frigid relationship between government and science. Other science-based non-governmental organizations and professional bodies should step up and advocate for basic science facilities, funds for salaries, local fabrication of equipment, and support for entrepreneurship. Affirmative action, childcare and other support for female scientists are also essential.

Scientists must step out of our campuses and laboratories more frequently to interact with the larger community. Last week, I joined some fellow academics for a conversation with local and federal politicians. I appeal to fellow scientists to make interactions like these a part of their work so that scientific issues are reckoned with, not just during election years but continually.

We need to harness our best minds to solve the mundane problems that have ground Nigeria to a near-halt. Science has to work for everyone in the country, so we don't continue wasting millions of hours simply waiting in line for basic needs, such as petrol.

Nature **614**, 595 (2023)

doi: https://doi.org/10.1038/d41586-023-00507-1

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00507-1

| Section menu | Main menu |

News in Focus

- **Disgraced CRISPR-baby scientist’s ‘publicity stunt’ frustrates researchers** [12 Feb 2023]
  • He Jiankui refused to answer researchers’ questions about his controversial 2018 experiments at weekend event.
- **WHO may soon end mpox emergency — but outbreaks rage in Africa** [10 Feb 2023]
  • African health officials worry that as the mpox outbreak wanes in wealthy countries, so too will the global interest in research and funding.
- **Researchers scramble as Twitter plans to end free data access** [14 Feb 2023]
  • A controversial policy change threatens to upend large social-media studies.
- **Marburg virus outbreak: researchers race to test vaccines** [15 Feb 2023]
  • Control measures such as quarantine could end the outbreak in Equatorial Guinea quickly — good news for inhabitants but a mixed blessing for clinical trials.
- **Outcry as scientists sanctioned for climate protest** [15 Feb 2023]
  • Researchers have asked the American Geophysical Union to reverse actions it took against scientists who demonstrated at a December meeting.
- **‘Mirror neurons’ fire up during mouse battles** [15 Feb 2023]
  • Brain cells are crucial for triggering fights — but also become active when mice merely observe fights.
- **High-altitude balloons: a scientists’ guide to what’s up there and why** [16 Feb 2023]
  • The US has shot down four flying objects over fears of spy balloons. But what else are high-altitude balloons used for?
- **The fight to keep Ukrainian science alive through a year of war** [22 Feb 2023]
  • Researchers say science is bleeding in Ukraine — but they are determined to sustain their work with help from other nations.

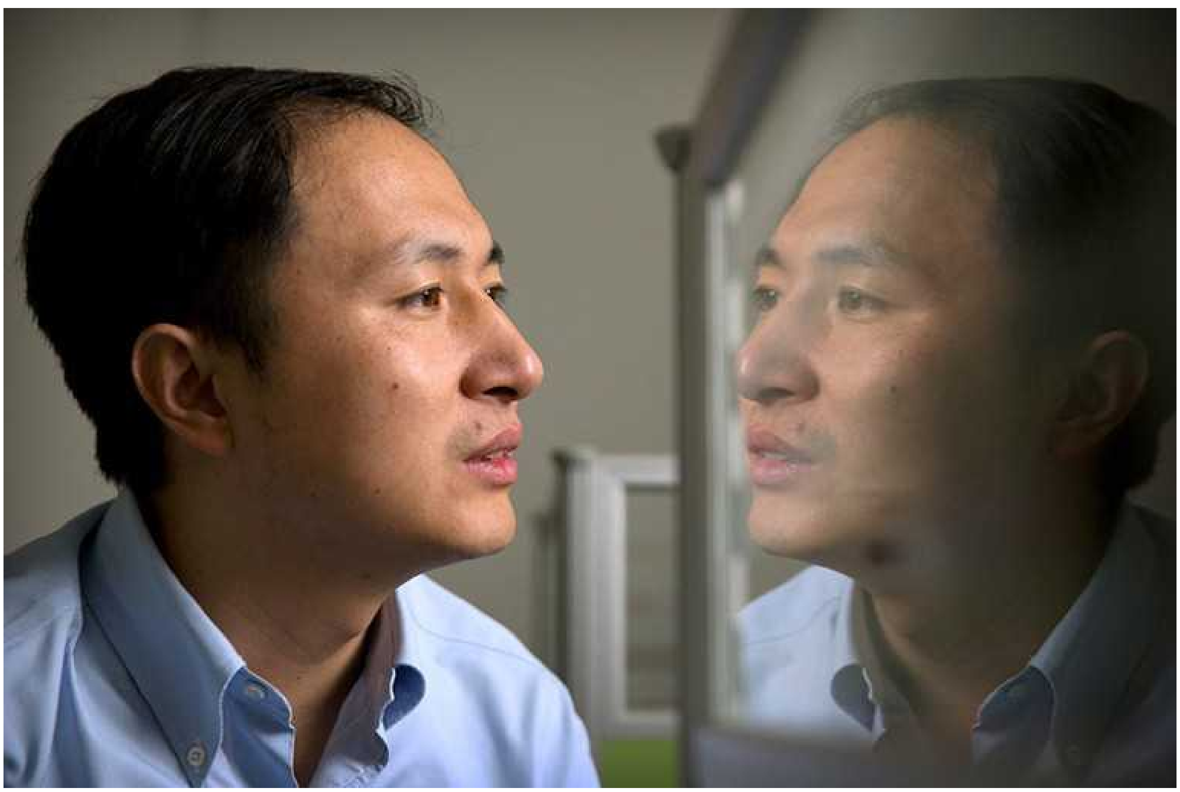
- **[Your brain could be controlling how sick you get — and how you recover]** [22 Feb 2023]
  • Scientists are deciphering how the brain choreographs immune responses, hoping to find treatments for a range of diseases.

- NEWS
- 12 February 2023
- Correction 13 February 2023

Disgraced CRISPR-baby scientist's 'publicity stunt' frustrates researchers

He Jiankui refused to answer researchers' questions about his controversial 2018 experiments at weekend event.

- Smriti Mallapaty



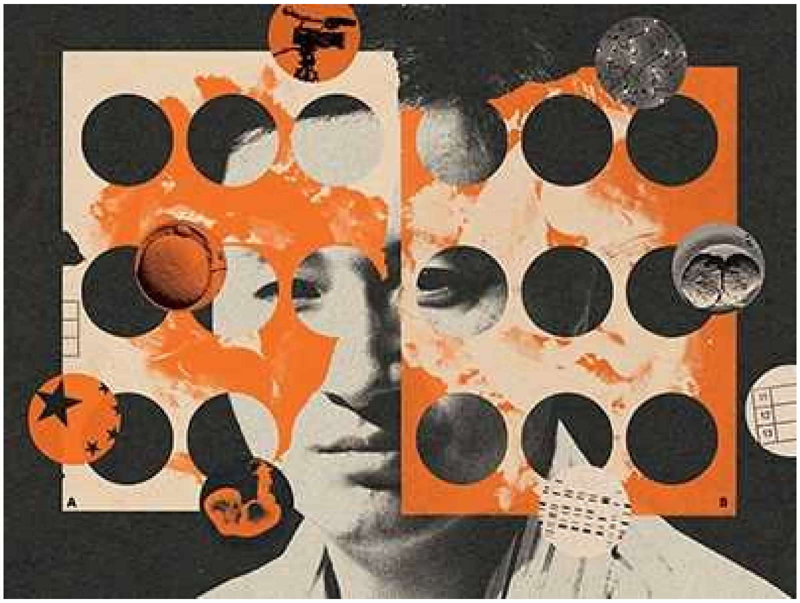
Biophysicist He Jiankui, who was released from prison last year, has declined to answer questions about his genome-editing experiments.Credit: Mark Schiefelbein/AP/Shutterstock

He Jiankui, the Chinese biophysicist who shocked the world with experiments that led to the first children with edited genomes, says research must accept moral and ethical constraints, but is otherwise refusing to speak about the work that sent him to jail for three years. His silence is frustrating some scientists, who say that he should answer questions about his past research before publicizing his latest plans to use genome-editing technology on people.

On Saturday, He spoke at a virtual and in-person bioethics event, hosted by the University of Kent in Cantebury, UK, that was promoted as “the first time that Dr. He has agreed to interact with Chinese bioethicists and other CRISPR scientists in a public event”. But during the talk, He did not discuss his past work and refused to answer questions from the audience, responding instead that questions should be sent to him by e-mail.

“This meeting has been very disappointing, notably the failure of He Jiankui to answer any questions,” says Robin Lovell-Badge, a developmental biologist at the Francis Crick Institute in London, who attended the event.

“A publicity stunt like today shows he doesn’t have much credibility, at least in the eyes of his peers,” says Eben Kirksey, a medical anthropologist at the University of Oxford, UK.



The CRISPR-baby scandal: what's next for human gene-editing

Hours before the event, He wrote in a tweet that he was not comfortable discussing his past work: "I feel that I am not ready to talk about my experience in past 3 years." He also said he would no longer be visiting the University of Oxford in March for a series of interviews with Kirksey, and that he would not be attending an international genome-editing summit at the Francis Crick Institute, where researchers will discuss the ethics of germline editing, also planned for March.

Kirksey declined to comment on the status of the Oxford interviews, but says He needs to clarify details about the circumstances surrounding his past experiments. In 2018, the world learnt that He had used CRISPR–Cas9 to edit a gene known as *CCR5* — which encodes an HIV co-receptor — with the goal of making the resulting children resistant to the virus. He implanted two embryos in one woman's uterus and one in another, resulting in twins and a third baby. The parents had agreed to the treatment because the fathers were living with HIV and the mothers were not.

He's experiments were widely condemned by scientists, and several scientific groups have since concluded that genome editing should not be used to make changes that can be passed on to future generations.

It is not known whether He’s previous work succeeded at protecting the children from HIV, or whether it caused any side effects. Without evidence of that, Kirksey remains sceptical about He’s future scientific plans.

He’s research plans

He was imprisoned for violating medical regulations in China. Since being released from prison last year, he has revealed on social media that he has set up a non-profit research laboratory in Beijing focused on developing affordable therapies for hereditary diseases such as Duchenne muscular dystrophy (DMD).

The weekend event was organized by the BioGovernance Commons initiative, a group that organizes monthly online meetings on ethical and regulatory issues between academics in China and those across Europe, North America and Asia. More than 80 researchers from 13 countries attended virtually, and He, together with some 20 academics and students, gathered at a venue in Wuhan, China.

He, who gave a 25-minute presentation in Chinese, with simultaneous English translation, briefly described his plans to develop a gene-editing drug for people living with DMD, for which he is currently raising funding from philanthropic organizations. He said he would not be accepting investments from commercial entities, to ensure that the therapies he develop are affordable, and that an international ethics committee would provide guidance on the work.

He also said that although genome editing in embryos destined for implantation is banned in China, the implementation of regulations on gene-editing technologies still lacks clarity. “Scientific research must be subject to constraints of ethics and morality,” he said at the end of his presentation. But He spent most of his talk describing the basics of genome-editing technology and its application in agriculture, infectious disease, diagnostics and human health.

“It bordered on being insulting to the conference organizers and the participants to have used up the time with details and information that were

not relevant,” says Françoise Baylis, a bioethicist at Dalhousie University in Halifax, Canada.

He’s presentation and failure to engage with scientists at the event shows he has not considered the social and ethical implications of his research, which suggests that he is not ready to work on genome editing, says Joy Zhang, a sociologist at the University of Kent and one of the event organizers. “We don’t want another scandal where desperate patients will be exploited for ventures on experimental or even illegal therapies,” says Zhang.

Anna Lisa Ahlers, a social scientist and China studies scholar at the Max Planck Institute for the History of Science in Berlin, commended He on agreeing to speak to his peers but says his presentation fell short of the scientists’ expectations. “From his talk, I would have gotten the impression that he is a salesman.”

Nature asked He to comment on criticisms about his talk from scientists. He responded by sharing a link to his earlier tweet.

Too much publicity?

Some researchers worry that interest in He Jiankui is diverting attention away from more important ethical issues around heritable genome editing. “This event puts the spotlight on He Jiankui — will he apologize? Is he displaying remorse?” says Marcy Darnovsky, a public-interest advocate on the social implications of human biotechnology, and executive director at the Center for Genetics and Society in Berkeley, California. Instead, she thinks researchers should focus on discussing whether there is a medical justification for heritable genome editing.

After He announced his results in 2018, the US National Academy of Medicine, the US National Academy of Sciences and the UK Royal Society concluded in a 2020 report that gene-editing technology was not ready for use in human embryos intended for implanting. And in July 2021, a committee convened by the World Health Organization advised against the use of heritable gene editing. Some 70 countries prohibit heritable genome editing, according to a 2020 policy review[1].

Nature **614**, 599-600 (2023)

doi: https://doi.org/10.1038/d41586-023-00382-w

Updates & Corrections

- **Correction 13 February 2023**: Due to a error in translation of He Jiankui's talk, a previous version of this story incorrectly stated that He said that germline editing should be banned. He actually said that germline editing is banned in China and elsewhere.

References

1. Baylis, F. *et al.* *CRISPR J.* **3,** 365–367 (2020).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00382-w

- NEWS
- 10 February 2023

WHO may soon end mpox emergency — but outbreaks rage in Africa

African health officials worry that as the mpox outbreak wanes in wealthy countries, so too will the global interest in research and funding.

- Max Kozlov



A doctor visits people with infections including mpox during his morning rounds at the Yakusu General Hospital in Tshopo, Democratic Republic of

the Congo.Credit: Arlette Bashizi/Reuters

Update: At a 15 February press conference, the WHO's director-general said that mpox would remain a public-health emergency. The agency's emergency committee will reconvene in three months to reassess the situation.

A World Health Organization (WHO) committee met earlier this week to decide whether the mpox outbreak — which began in May 2022 — is still a global public-health emergency, and the agency could soon declare it over. The outbreak has subsided in countries including the United Kingdom and the United States, thanks to the deployment of vaccines and therapeutics, as well as changes in awareness and social behaviour. But the same is not true in some nations in West and Central Africa, which have been battling the monkeypox virus for decades, and where the disease's toll has historically been highest.



Monkeypox in Africa: the science the world ignored

Even while global mpox infections have plummeted from more than 1,000 new infections a day at their peak to fewer than 70, numbers have not fallen substantially in African countries. The Democratic Republic of the Congo (DRC), for instance, has reported more than 1,000 suspected cases since October. But even those numbers are a "gross underestimation" of reality,

says Dimie Ogoina, an infectious-disease physician at Niger Delta University in Amassoma, Nigeria. Many infections are never confirmed because of an underfunded testing and virus-surveillance system.

African researchers commend the increased attention paid to the long-neglected virus during the global outbreak. But they say that wealthy countries were the main beneficiaries. Vaccines and treatments have not materialized in Africa, and the knowledge gained about the virus has done little to change the course of the outbreak on the continent.

"The majority of the published literature of monkeypox is from the global north," Ogoina adds. "We are still working largely blindly here" to understand transmission patterns and effective responses in Africa, he says.

A long battle

Before 2022, almost all of the world's human cases of mpox had been in Africa. The disease (which is now called mpox, although the virus continues to go by monkeypox) causes fever and painful fluid-filled lesions on the face, hands and feet — and in severe cases, hospitalization and death. Some African nations have been grappling with mpox outbreaks ever since scientists identified the first human infection in the DRC in 1970.



Dimie Ogoina, an infectious-disease physician at Niger Delta University in Amassoma, Nigeria, has been studying mpox for years.Credit: Emeke Obanor

Outside Africa over the past year, the virus has spread mainly among men who have sex with men, through close skin-to-skin contact. Transmission patterns and clinical manifestations of the disease are much harder to pin down in some African countries, where rodent species are thought to naturally harbour the virus and regularly spread it to humans. For example, whereas more than 95% of people infected with the virus in the United States have been men, only about 60% of those who have had confirmed infections in Nigeria are men.

And infections in Africa are still flying under the radar, Ogoina says. Out of more than 7,200 infections on the continent suspected of being mpox in 2022, only about 1,200 could be confirmed by standard diagnostic methods, because of an underfunded testing and surveillance infrastructure. The DRC has reported more than 70% of the continent's suspected cases — of which only 279 have been confirmed.

Treatments out of reach

The millions of smallpox vaccines purchased by and distributed throughout wealthy countries such as Canada and the United Kingdom have also largely been out of reach for fighting mpox in African nations. (The viruses that cause smallpox and mpox are related, so health officials hoped that smallpox vaccines would also work for mpox.) The African Union will receive a 50,000-dose donation from South Korea — the continent's very first vaccines not designated for research purposes — according to a 29 November announcement by the Africa Centres for Disease Control and Prevention. The agency, however, did not respond to *Nature*'s queries as to when those doses would be delivered, or to which countries.



What does the future look like for monkeypox?

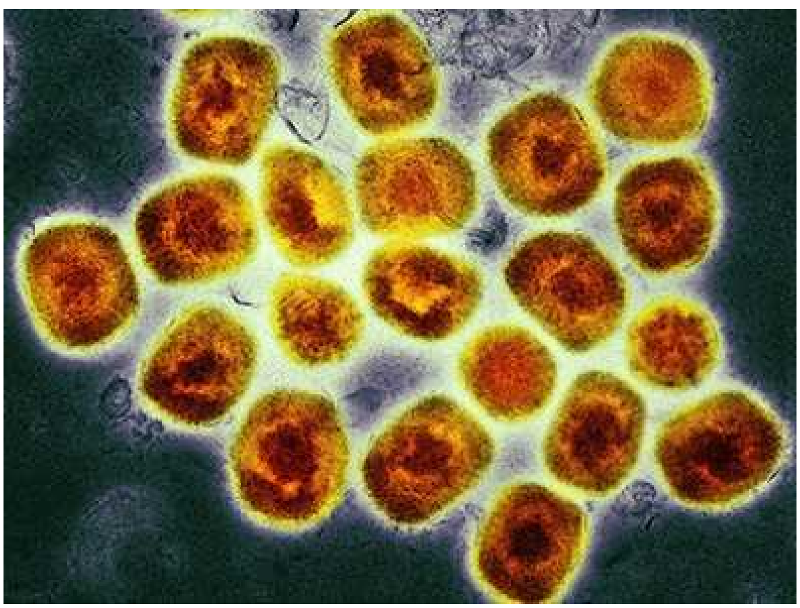
Rosamund Lewis, technical lead for mpox at the WHO, based in Geneva, Switzerland, acknowledges that efforts to provide vaccines and treatments to African nations have "not been moving fast enough for anyone's satisfaction". Supplies and donations have been scarce; in the meantime, the WHO's regulatory department is still reviewing data collected from other countries about the vaccines' safety and efficacy, she says.

There has been a similar delay in African nations accessing tecovirimat, an antiviral approved for smallpox that is thought to work against mpox. US regulators authorized the drug for compassionate use against mpox during

the global outbreak, and more than 6,800 doses have been administered there. By contrast, people in Africa have access to tecovirimat only through clinical trials, which are now starting up in several African countries. Piero Olliaro, a specialist in poverty-related infectious disease at the University of Oxford, UK, has been conducting a small trial of the drug in the Central African Republic, in which 20 people have received it. “That’s less than a drop in the ocean,” he says. “How come we offer tecovirimat to all mpox patients in the US, where the drug is registered only for smallpox, but need a [randomized, controlled trial] to prove it works in Africa?”

‘Complacency isn’t an option’

Still, the global focus on the virus has changed a few things in Africa. For one, because of all the published case reports, physicians who had never heard of the disease now know what the telltale lesions look like, Ogoina says.



The monkeypox virus is mutating. Are scientists worried?

And the global interest has inspired some research projects in Africa. For example, Rosemary Audu, a virologist and director of research at the Nigerian Institute of Medical Research in Lagos, and her colleagues created the country’s first mpox test for research use, to complement the ones that

health officials use to diagnose infections. Audu and her colleagues also received a Can$3-million (US$2.25-million) grant from Canada's International Development Research Centre and the Canadian Institutes of Health Research, both based in Ottawa, that she hopes to use to better understand the virus's transmission and spread in Nigeria. As part of this project, she plans to apply her team's mpox test to samples stored at clinics around the country that treat sexually transmitted infections, to look for undiagnosed cases.

Because the monkeypox virus continues to cause deaths in Africa and could contribute to future outbreaks everywhere, complacency isn't an option, Ollario says. "Monkeypox may come back again in a more nasty way," he adds. In particular, researchers worry that a strain of the virus found in Central Africa, called clade I, could spread. This one kills about 10% of infected people, making it more lethal than the strain circulating in West Africa that is thought to have sparked the global outbreak. There have already been signs that the clade I strain can spread beyond its typical range: it was responsible for an outbreak last year at a refugee camp in Sudan that led to more than 15 confirmed and 180 suspected infections, Lewis says.

Nature **614**, 600-601 (2023)

doi: https://doi.org/10.1038/d41586-023-00391-9

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00391-9

- NEWS
- 14 February 2023

Researchers scramble as Twitter plans to end free data access

A controversial policy change threatens to upend large social-media studies.

- Heidi Ledford



Twitter announced on 2 February that it would end free access to its API.Credit: Tayfun Coskun/Anadolu Agency via Getty

Akin Ünver has been using Twitter data for years. He investigates some of the biggest issues in social science, including political polarization, fake

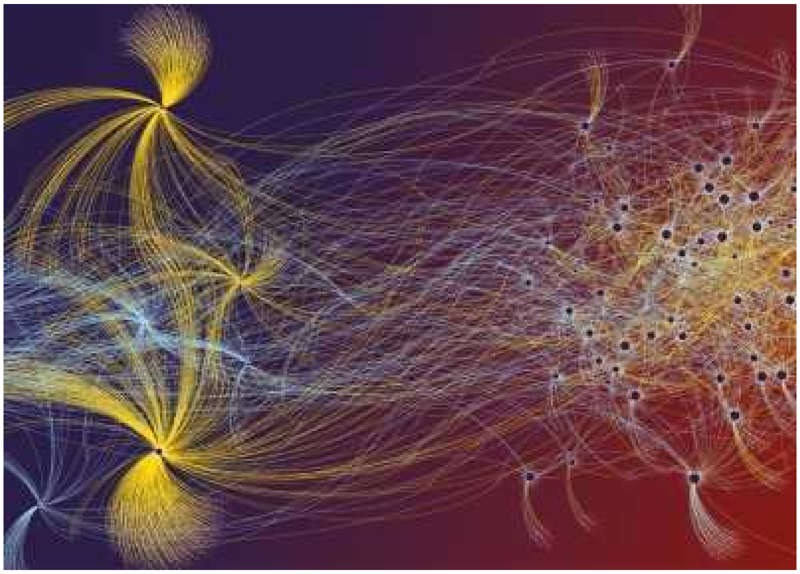
news and online extremism. But earlier this month, he had to set aside time to focus on a pressing emergency: helping relief efforts in Turkey and Syria after the devastating earthquake on 6 February.

Aid workers in the region have been racing to rescue people trapped by debris and to provide health care and supplies to those displaced by the tragedy. Twitter has been invaluable for collecting real-time data and generating crucial maps to direct the response, says Ünver, a computational social scientist at Özyeğin University in Istanbul.

So when he heard that Twitter was about to end its policy of providing free access to its application programming interface (API) — a pivotal set of rules that allows people to extract and process large amounts of data from the platform — he was dismayed. “Couldn’t come at a worse time,” he tweeted. “Most analysts and programmers that are building apps and functions for Turkey earthquake aid and relief, and are literally saving lives, are reliant on Twitter API.”

The end of free access

Twitter has long offered academics free access to its API, an unusual approach that has been instrumental in the rise of computational approaches to studying social media. So when the company announced on 2 February that it would end that free access in a matter of days, it sent the field into a tailspin. “Thousands of research projects running over more than a decade would not be possible if the API wasn’t free,” says Patty Kostkova, who specializes in digital health studies at University College London.



[How Facebook, Twitter and other data troves are revolutionizing social science]

Since then, researchers have been anxiously awaiting further details of Twitter's plans. The date that free access is due to end has been pushed back twice: first on 8 February, and then in an announcement on 13 February that said the new platform would be unveiled in "a few more days". Furthermore, key information about a proposed new payment plan is missing, says David Lazer, a political scientist at Northeastern University in Boston, Massachusetts. The company says that it will offer tiered pricing, with the cheapest version costing US$100 a month for "low-level" access. But it hasn't said what that will entail. "People can have different orders of magnitude in their imagination as to what 'low-level' means," says Lazer. "If they're optimizing the pricing for corporate customers, there's no way academics are going to be able to afford it."

The rapid change in policy will cause difficulties for some European Union-funded projects, whose proposals were written with the assumption that API access would remain free, says Ünver. And prices could quickly mount if each researcher on a project had to pay the monthly fee.

Even if well-funded researchers find a way to get the data they need, the policy change will exacerbate inequalities for students or researchers from low-income countries, for whom the monthly fees could be too high to sustain long-term studies. "Twitter is a global platform, and this decision has

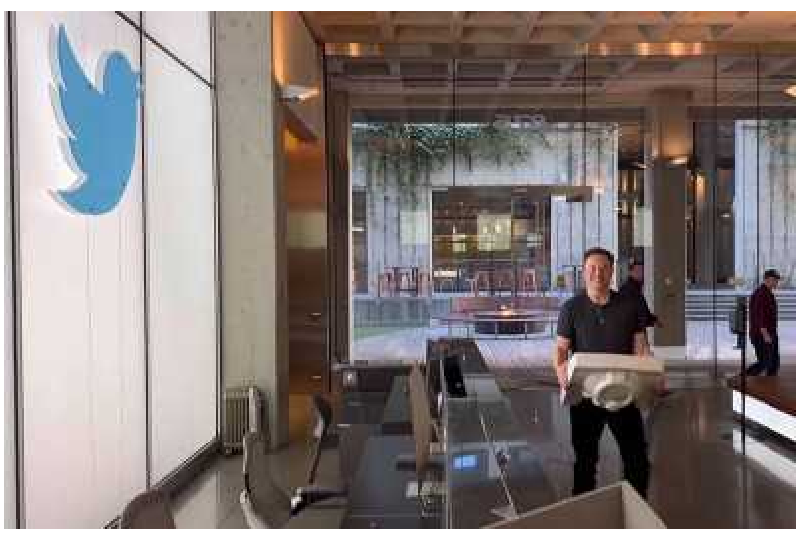
global ramifications," says Renee DiResta, research manager of the Stanford Internet Observatory at Stanford University in California. "Discontinuing free access will break free tools developed to democratize research."

The move also runs counter to the spirit of the EU's Digital Services Act, which came into force last November, says Claes de Vreese, who studies political communication at the University of Amsterdam. As a very large social-media platform, Twitter could eventually be compelled under the legislation to ensure data transparency and access for researchers.

New problems

When the API announcement came, researchers were already feeling unmoored as a result of changes at Twitter after billionaire Elon Musk bought the company last October. Ünver says that since then, his data requests using the API have yielded some worrisome glitches. For example, a program that uses earthquake-related keywords in Turkish doesn't always return all of the tweets that he sees on his own Twitter feed — even if he searches with the exact words of the tweet. And a few times, he has spotted duplicate tweets in his data. "Both problems are new," he says.

Restricting access could create further data difficulties. De Vreese says that his colleagues have been frantically collecting data since word of the API change broke. Over time, he says, limiting access to new information could render the field over-reliant on previously downloaded data sets.



[How Musk’s takeover might change Twitter: what researchers think]

Before the Twitter announcement, Lazer’s lab had gathered enough data to crank out a few academic papers, including one that looks at the impact of Twitter’s interventions, such as banning certain users, after a mob attacked the US Capitol on 6 January 2021. “In the longer run, none of that work can be replicated,” he says. And if another insurrection occurs, researchers will not be able to study it in the same way.

Some projects might simply grind to a halt, because other social-media platforms do not offer the same degree of access as Twitter, or the same quality of data. Kostkova and her colleagues have been searching for ways to keep working on their epidemic-intelligence research — in which they have used Twitter for more than a decade to study subjects including swine influenza, vaccine hesitancy and COVID-19. “At the moment there is nothing,” she says.

The situation also highlights how the availability of Twitter’s data has led the field to depend on the platform, says de Vreese. “This is a moment in time when we have to rethink,” he says. “It’s reigniting a conversation that we know we should be having: that there is too much Twitter-reliance in research.”

Nature **614**, 602-603 (2023)

doi: https://doi.org/10.1038/d41586-023-00460-z

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00460-z

- NEWS
- 15 February 2023
- Correction 17 February 2023

Marburg virus outbreak: researchers race to test vaccines

Control measures such as quarantine could end the outbreak in Equatorial Guinea quickly — good news for inhabitants but a mixed blessing for clinical trials.

- Ewen Callaway



Marburg virus particles in infected tissue. The virus causes a deadly disease characterized by haemorrhagic fever.Credit: AMI Images/Science Photo Library

Health officials worldwide are sprinting to test whether experimental vaccines can protect against a deadly illness, after Equatorial Guinea confirmed its first outbreak of Marburg virus disease on 13 February. The virus is related to Ebola, and causes similar symptoms of haemorrhagic fever. It has a fatality rate of up to 88%.

The World Health Organization (WHO) in Geneva, Switzerland, convened an urgent meeting yesterday to discuss the feasibility of testing Marburg vaccines that are in various stages of development. But the odds are against a successful trial, they say, because other control measures such as quarantine could end the outbreak before a single vaccine dose can be administered.

“I cannot emphasize enough the need for speed,” said John Edmunds, an epidemiologist at the London School of Hygiene & Tropical Medicine, at the WHO meeting.



The radical plan for vaccine equity

The outbreak is in the north of Equatorial Guinea, in the Kié-Ntem province, which borders Cameroon and Gabon. It has been linked to 9 deaths among 25 suspected cases, with the first known case dating to early January. This makes it larger than many of the 16 Marburg outbreaks that have previously been detected, Edmunds tells *Nature*. “Outbreaks have tended to be small and finish relatively quickly after effective interventions have been put in place.”

The exceptions are a 1998–2000 outbreak in the Democratic Republic of the Congo that was linked to 154 cases and 128 deaths, and a 2004–05 epidemic in Angola that caused 227 deaths among 252 reported cases.

Challenging logistics

At this week's WHO meeting, officials discussed the practicalities of testing Marburg virus vaccines in Equatorial Guinea. All the leading contenders are viral-vector vaccines, similar to the COVID-19 vaccine developed by AstraZeneca and the University of Oxford, UK.

The Sabin Vaccine Institute in Washington DC has a candidate vaccine that uses a modified chimpanzee adenovirus to deliver instructions for cells to make a Marburg virus protein, whereas a candidate made by Janssen in Beerse, Belgium, uses the human adenovirus on which the company's successful COVID-19 vaccine was based (Janssen is a subsidiary of Johnson and Johnson).

Candidates from Public Health Vaccines (PHV) in Cambridge, Massachusetts, the International Aids Vaccine Initiative (IAVI) in New York City and Auro Vaccines in Pearl River, New York, are based on weakened forms of vesicular stomatitis virus — the vector that is used in the first approved Ebola vaccine.



The slow roll-out of the world's first malaria vaccine

None of the vaccines is available in large quantities, developers said at the meeting: availability ranges from a few hundred doses in the case of the Sabin and PHV vaccines to a few thousand for Janssen's candidate. IAVI has no available doses of its vaccine. Only the Janssen and Sabin vaccines have been tested in humans, in early-stage trials in the United States. Monkey studies suggest that all the leading candidates offer strong protection against Marburg virus disease.

If a vaccine trial in Equatorial Guinea were to go ahead, an independent group of experts that advises the WHO would make decisions about which vaccines to test, says Ana Maria Henao-Restrapo, who co-leads the WHO's R&D Blueprint effort to lay the groundwork for such studies during outbreaks. Any trial would also require the permission and involvement of Equatorial Guinea's government.

Even if a trial can get off the ground, it's unlikely that enough cases will develop before the current outbreak comes under control for researchers to determine conclusively whether any vaccine is effective or not, Edmunds says. "It's a double-edged sword, isn't it? It's good news for public health and the people of Equatorial Guinea, but perhaps bad news for science."

But evidence pointing to the effectiveness of any vaccine could be gathered across multiple outbreaks, researchers at the meeting said. A vaccine trial in Equatorial Guinea could also provide valuable data on the safety of vaccines and the immune response they generate in populations at risk of future outbreaks.

Nature **614**, 603 (2023)

doi: https://doi.org/10.1038/d41586-023-00468-5

Updates & Corrections

- **Correction 17 February 2023**: An earlier version of this article stated incorrectly that IAVI's Marburg vaccine was being co-developed with Merck.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00468-5

- NEWS
- 15 February 2023

Outcry as scientists sanctioned for climate protest

Researchers have asked the American Geophysical Union to reverse actions it took against scientists who demonstrated at a December meeting.

- Myriam Vidal Valero



Researchers have increasingly taken up activism in the face of the climate crisis.Credit: Christian Mang/Reuters

Update: After concluding an ethics investigation, on 17 February the AGU issued a statement that it will restore the abstracts of Abramoff and Kalmus to its December meeting programme, and that it aims to clarify its policies and codes of conduct moving forward. The society also invited Abramoff and Kalmus to work with it to "expand existing methods of climate activism and engagement through AGU".

More than 2,000 researchers from around the world have signed a letter asking the American Geophysical Union (AGU) to reverse actions it took against two scientists who briefly protested at its annual meeting in Chicago, Illinois, in December. The scientists had urged their fellow attendees to speak out about the climate crisis. The demonstration was minor, the letter says, and it is scientists' responsibility to speak up about global warming.



Young people's climate anxiety revealed in landmark survey

The episode resurfaces a long-standing debate about whether climate scientists should engage in activism — and raises new questions about whether their careers will be threatened if they do.

On 15 December 2022, Rose Abramoff, an Earth scientist then at Oak Ridge National Laboratory in Tennessee, and Peter Kalmus, a climate scientist at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California, walked on stage at the beginning of an AGU plenary talk and held up a banner reading

“Out of the lab & into the streets”. Moments later, the organizers of the session took away the sign and asked them to leave, which they did.

In response to the protest, the AGU removed the scientists’ abstracts from the meeting programme, expelled them from the meeting and opened cases of professional misconduct against them. The cases are still ongoing. In January, Abramoff was fired from Oak Ridge as a result of the incident. Kalmus still works at JPL, but declined to tell *Nature* whether officials there had reprimanded him.

Climate scientists are becoming more and more frustrated by what they view as a lack of action from governments and institutions to curb emissions, so it makes sense that they would choose activism, says Almut Arneth, a biologist at the Karlsruhe Institute of Technology in Germany who signed the letter. The AGU’s response is disproportionate to a non-violent protest, she adds. “We need to make sure that we as scientists are stepping up to stop something that could potentially be quite dangerous for free speech in science.”



Kalmus and Abramoff protest on stage at the AGU meeting in December.Credit: Dwight Owens

In particular, many younger researchers see themselves as being most at risk from climate change, and so are increasingly turning to activism. The society's actions in this case set a precedent that could be especially damaging for early-career scientists such as Abramoff, and other researchers in positions without substantial power, says Ana Bastos, a geophysicist at the Max Planck Institute for Biochemistry in Jena, Germany, who signed the letter. "They might fear for their careers, for their jobs, for their future if they want to become politically engaged."

The AGU, which has 60,000 members who include Earth and space scientists, declined to answer *Nature*'s questions because of the ongoing investigations. But a spokesperson said that the society's ethics policy and meetings code of conduct require "attendees to treat everyone with respect, and this includes respecting presenters' time to speak and audiences' time to listen".

The spokesperson also said that the society provides opportunities for people "year round" to reserve space for public engagement at all its meetings, and has a proactive stance on "aggressively" addressing the urgency of climate change.

Civil disobedience

This wasn't the first time that Abramoff and Kalmus had protested to demand action on climate change. Both had even been arrested for civil disobedience. For instance, they were detained last November after chaining themselves to the fence of Charlotte Douglas International Airport in North Carolina to protest against emissions from private jets, as part of global demonstrations organized by the activist group Scientist Rebellion.



Scientist Rebellion: researchers join protesters at COP26

Because the climate crisis is so dire, Abramoff says, she is willing to be arrested or fired to fight it — but she wasn't expecting that to happen as a result of the AGU meeting. "I was surprised. And I was sad, because I really enjoyed working at the laboratory."

According to a 10 January opinion piece that Abramoff wrote for *The New York Times*, Oak Ridge fired her because, it said, she had misused government resources by engaging in a personal activity on a work trip, and she hadn't adhered to its code of business ethics and conduct.

A spokesperson for Oak Ridge declined to answer *Nature*'s queries about why it had fired Abramoff, except to say that the lab respects the rights of all employees to pursue their personal interests outside work.

Impartial observers no more

Nature polled more than 200 climate scientists in 2021 about activism and other issues. One-quarter said they had participated in climate demonstrations, and two-thirds said they engage in climate advocacy. Aaron Thierry, an ecologist at Cardiff University, UK, says that evidence about

global warming has been ignored for decades, so civil disobedience is a natural next step for scientists seeking to effect change.



Top climate scientists are sceptical that nations will rein in global warming

In the past, scientists have been reluctant to raise their voices in case activism compromised their image as unbiased gatherers of data and called into question the legitimacy of their work, Thierry says. But climate scientists can no longer be impartial observers, he adds.

“We are going to be affected by the consequences of climate change as citizens and as individuals,” he says. Taking action “doesn’t make us bad scientists, it doesn’t undermine the work that we do. If anything, I think, it makes us better scientists, because it means that we’re engaging with our work and what it means to society.” To Thierry, the AGU’s response is terrifying because it’s a step backwards from what the climate crisis requires.

Making room for debate

As a society that represents a large portion of the climate-change-research community, Abramoff says, the AGU could have used the protest as an opportunity to embrace the debate over activism and make room to discuss

it. Organizers could have said, “Well, you know, we’ve got a situation here, we don’t know how to deal with it — why not set aside a panel discussion at the next meeting?” she says.



Why young climate activists have captured the world’s attention

Some policymakers are asking similar questions. In response to the sanctions against Abramoff and Kalmus, US senator Edward Markey (Democrat, Massachusetts) wrote to the AGU asking for an explanation. The society awards a monetary prize to scientists who engage in climate communication, he wrote. So “it is as baffling as it is disappointing that AGU apparently paired its important efforts to promote global understanding of climate change with efforts to suppress actions taken in furtherance of it”.

The AGU says it will issue a report on its investigations concerning Abramoff and Kalmus on 17 February.

Nature **614**, 604-605 (2023)

doi: https://doi.org/10.1038/d41586-023-00440-3

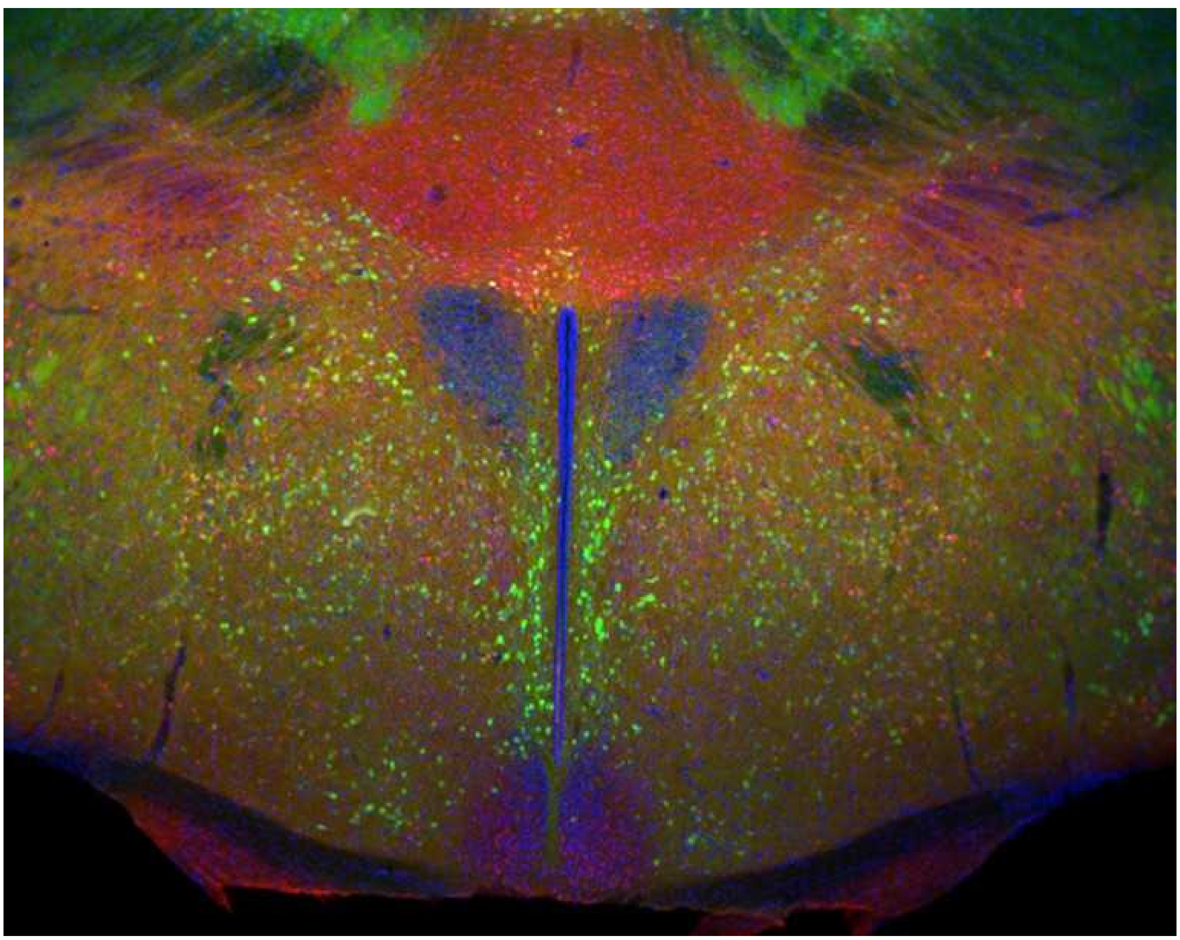
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00440-3

- NEWS
- 15 February 2023

‘Mirror neurons’ fire up during mouse battles

Brain cells are crucial for triggering fights — but also become active when mice merely observe fights.

- Max Kozlov



The hypothalamus of a mouse (red marks a signalling protein). This brain region contains cells called mirror neurons that are involved in aggressive behaviour.Credit: NIH/Avalon

A group of brain cells in mice becomes active both when the animals fight and when they watch other mice fight, a study[1] shows. The work hints that such 'mirror neurons', which fire when an animal either observes or takes part in a particular activity, could shape complex social behaviours, such as aggression.

The mirror neurons described in the study are the first to be found in the hypothalamus, an evolutionarily ancient brain region — suggesting that mirror neurons' original purpose might have been to enhance defence and, ultimately, reproductive success, the authors speculate. The study was published in *Cell* on 15 February.

"We've now shown that mirror neurons functionally participate in the behaviours they're mirroring," says Nirao Shah, a neuroscientist at Stanford University in California who co-authored the study. "That changes what we think about mirror neurons."

Monkey business

First identified in monkeys in the 1990s, mirror neurons generally fire when an animal takes a certain action, but they also fire when it sees another animal perform the same action. Previous work has linked mirror neurons' activity to simple behaviours, such as reaching for an object, but not to complex social behaviours, such as fighting.

But exactly how mirror-neuron activity contributes to cognitive functions has been controversial, says Pier Francesco Ferrari, a neuroethologist at the Institute of Cognitive Science Marc Jeannerod in Lyon, France. Some researchers have argued that the fact that mirror neurons fire both when an animal observes a behaviour and when it performs that behaviour itself shows that these neurons are involved in a higher-order awareness of others' actions — and perhaps even contribute to empathy. But others say that there is little evidence to support this theory.

Mirror neurons have mostly been found in the prefrontal cortex, an area of the brain responsible for regulating social behaviours. But experiments by Taehong Yang, a neuroscientist at Stanford University who co-authored the latest study, and his colleagues made them wonder whether there might be mirror neurons in the ventromedial hypothalamus (VMH), a region that some researchers have called the brain's attack centre. In 2017, Yang and his colleagues found[2] that in certain social situations, activating the VMH neurons in mice did not trigger aggression. This made the researchers think that the VMH might include mirror neurons, which are sensitive to other individuals' actions.

Full of fight

Yang and his colleagues recorded VMH neurons when male mice watched others fight. They found that some VMH neurons fired both when an animal itself acted aggressively and when it watched others do so. Using a genetic-marker system, the researchers also showed that these neurons don't just act as bystanders to the fight but are essential for triggering fighting.

Ferrari says that the study doesn't answer why animals have mirror neurons. But he says that the experiments are well done and that the results hint that mirror neurons might be essential to a social behaviour such as aggression. For example, the mirror neurons — or an entire system of neurons of which VMH neurons are part — of a mouse watching a fight might encode which animal wins that fight.

Scientists knew previously that neurons in the hypothalamus control behaviour, but this is the first evidence that neurons in this region "are actually able to sense or respond to another animal's social experience", says Dayu Lin, a social neuroscientist at NYU Langone Health in New York City. It is especially striking that these neurons respond to visual cues in an animal species that relies so heavily on smell, she adds.

Nature **614**, 605 (2023)

doi: https://doi.org/10.1038/d41586-023-00418-1

References

1. Yang, T. *et al.* *Cell* https://doi.org/10.1016/j.cell.2023.01.022 (2023).

2. Yang, T. *et al.* *Neuron* **95**, 955–970 (2017).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00418-1

- NEWS EXPLAINER
- 16 February 2023

High-altitude balloons: a scientists' guide to what's up there and why

The US has shot down four flying objects over fears of spy balloons. But what else are high-altitude balloons used for?

- Nicola Jones



High-altitude balloons are used for science, for example to get a clear view of space, or to test instruments.Credit: NG Images/Alamy

When the US government shot down a suspected Chinese surveillance balloon off the South Carolina coast on 4 February, it ramped up geopolitical tensions with China and prompted officials to tighten their radar search for other high-flying unidentified objects. Since then, the United States has shot down three more objects — which it now seems were probably not spy balloons. How many balloons are in the skies, what are they used for, and what will the recent incidents mean for balloon users?

The suspected spy balloon was 60 metres tall, carrying a payload weighing around one tonne. The US state department says the payload carried devices to intercept sensitive communications; China says that it was a civilian research airship gone astray.

The other objects have been smaller and less well-described by officials: an object "about the size of a small car" over Alaska; a small cylinder over the Yukon in Canada; and an octagonal structure with strings over Lake Huron in Michigan. All flew in commercial airspace at an altitude of around 6–12 kilometres. The US government now says the "leading explanation" is that they "could just be balloons tied to some commercial or benign purpose".

The United States has recently started taking 'unidentified anomalous phenomena' (UAPs) — including flying objects — more seriously. Last year, NASA established a team of scientists, technology, flight and space experts to investigate UAPs, citing both national security and air safety. And in January, the US Office of the Director of National Intelligence revealed a dramatic uptick in UAP reporting. For the 17 years before March 2021, the agency catalogued 263 UAP reports; since that date, there have been 247 in less than 2 years. Out of 366 reports analysed, 163 were characterized as balloons, 26 as uncrewed aircraft and 6 as clutter.

Do look up

Balloons are a valuable tool for getting a good view downwards or upwards, from an altitude higher than drones or planes can reach and at a lower cost than satellites.

By far the majority are weather balloons: these are launched twice a day simultaneously from almost 900 locations worldwide, according to the US National Weather Service. They transmit data about temperature, humidity, pressure and location, and are disposable. The thin balloons — typically made of biodegradable latex — expand at altitude to about 6 metres in diameter. Flights are designed to go straight up to about 30 kilometres, and last for only a few hours.

Some scientists use much larger, longer-lasting balloons, for example to get a clear view of space, or to test instruments destined for high altitudes. NASA's Wallops Flight Facility in Virginia manages the launch of about 10–15 scientific balloons each year worldwide. These can carry around 3,000 kilograms, expand to be larger than a football stadium and fly to an altitude of 37 kilometres.

Other balloon users include science students, companies and amateur enthusiasts. Jason Krueger says his company StratoStar in Fishers, Indiana, has helped students and companies to launch more than 1,000 high-altitude balloon missions since 2006. Student projects have included investigating whether Post-it notes are still sticky after a flight to near-space and the impacts of high-altitude radiation on blood samples.

Corporate uses of balloons include providing Wi-Fi in remote regions, and some amateur enthusiasts launch picoballoons. These silver-coloured, plastic Mylar balloons typically measure less than 1 metre in size and are harmless, says Krueger, carrying amateur radios and payloads of just a few grams. But these shiny balloons would "light up radar like nobody's business", says Krueger, and their typical flight altitude is around 12 kilometres.

Blame game

Many balloons can be discounted from the list of those shot down: weather balloons make short flights and don't drift at 12-kilometre altitudes, for example. But that still leaves plenty unaccounted for worldwide. "There are flights every day of research, corporate and hobbyist balloons," says Robert Rohde, a scientist at the environmental non-profit organization Berkeley

Earth, who lives in Zurich, Switzerland. “I suspect that what they shot down are related to one of those categories.”

The US Federal Aviation Administration doesn’t require tracking devices for payloads under 5.4 kilograms, or for launches or flight paths for such loads to be declared. But even small packages can use large balloons. If such objects start to attract military attention, perhaps they, too, should be tracked, says Rohde. “I don’t feel like it’s necessary from a safety point of view, but if there’s a legitimate concern about small balloons from other states, we should probably make sure these things are identified.”

Krueger doesn’t think that’s needed. Instead, he says, the US government should “get better at assessing what is a threat”.

Nature **614**, 606 (2023)

doi: https://doi.org/10.1038/d41586-023-00482-7

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00482-7

- NEWS FEATURE
- 22 February 2023

The fight to keep Ukrainian science alive through a year of war

Researchers say science is bleeding in Ukraine — but they are determined to sustain their work with help from other nations.

- Aisling Irwin



Physicist Kseniia Minakova’s laboratory was destroyed after a Russian missile hit the National Technical University Kharkiv Polytechnic Institute.

Credit: Mykhailo Kirichenko

A year into Russia’s war on Ukraine, Kseniia Minakova is rebuilding her research laboratory and her career.

Minakova, a 34-year-old physicist in Kharkiv, remained in Ukraine when Russia invaded last February. Despite the crisis, on 19 August, she was buoyed by clinching a collaboration with US scientists on solar-energy research. But that same night brought disaster. A Russian missile hit her campus, killing a guard and destroying her optics and photonics laboratory, at the National Technical University Kharkiv Polytechnic Institute.



Kseniia Minakova salvaged equipment from the rubble of her laboratory.
Credit: Vasyl Holosniy

Minakova and her colleagues salvaged microscopes, welding equipment and computers from the rubble and moved into a smaller space nearby. Her US partners at Tulane University in New Orleans, Louisiana — one of whom, Denys Bondar, is Ukrainian-born — sent a support package, including solar cells, a thermal imaging camera and other equipment to Kharkiv. Minakova

now has a lab space one- quarter the size of what she lost, but the joint work with Tulane, part of a wider collaboration between the two universities, is under way.

Despite daily, three-hour power outages and frequent air-raid alarms, Minakova is continuing her research into high-efficiency solar cells, and she hopes to visit Tulane for training later this year. She delivers lectures and lab-demonstration classes online to her undergraduate students, who aren't allowed on Kharkiv's campus for safety reasons. In her spare time, she volunteers to help older residents find water and medicine, and helps children to embroider goodwill tokens for soldiers. "The best idea for living is to have no free time," she says.



Kseniia Minakova in 2021, before the war.Credit: Valery Tayemnytskyi

Minakova is one of tens of thousands of scientists in Ukraine striving to keep their research going as their country enters a second year of resisting Russia's aggression.

Most of the country's researchers — who totalled 60,000 before the war, with some 35,000 scientific support staff — have remained in Ukraine. Estimates of the number who have left vary widely, but the Ministry of Education and Science of Ukraine thinks some 6,000 scientists are currently out of the country because of the war. Of those still in Ukraine, some have

chosen to fight on the front line — and some have died there; others have lost family members. Many are internally displaced because their workplace is damaged, destroyed or impossible to operate in owing to a lack of power. Some have lost jobs, project funds or a proportion of their salary, as money is repurposed for the war effort.

As of January, 91 research and higher education institutes had been damaged; with 4 totally destroyed, according to the science ministry. Some 228 remain unharmed. The renowned Kharkiv Institute of Physics and Technology and the world's largest decametre-wavelength radio telescope, which belongs to the Institute of Radio Astronomy in Kharkiv, were heavily damaged.



The National Technical University Kharkiv Polytechnic Institute, after buildings on campus were hit by a Russian missile in August.Credit: Sofiia Bobok/Anadolu Agency via Getty

In December last year, the National Academy of Sciences of Ukraine accused Russia of "purposefully destroying science in Ukraine as a profession", an act it called "scienticide".

Science in Ukraine is “bleeding”, says Igor Komarov, a 58-year-old organic chemist and the director of the Institute of High Technologies at the Taras Shevchenko National University of Kyiv. Many of the scientists who remain, however, are determined to toil on against the odds. And increasingly, the international community is finding ways to help.



Organic chemist Igor Komarov says science in Ukraine is “bleeding”.Credit: Igor V. Komarov

Although foreign institutions have been offering support to Ukrainian scientists who have left the country since the early days of the invasion, they are now tailoring that help so that these scientists can maintain their ties with institutions at home. The foreign institutions are also increasingly targeting funds towards researchers inside the country — for salaries, equipment and infrastructure repairs, as well as for virtual collaboration efforts. In the middle of the tragedies of lost lives, bombed buildings and thwarted careers, many Ukrainian scientists are forging collaborations with institutions outside the country that, some think, are laying a strong foundation for a thriving, Western-facing scientific community after the war.

Rolling blackouts

Ukraine has dealt with Russia’s unrelenting destruction of energy infrastructure by putting the country on a schedule of rolling blackouts. Water supply and heating are also subject to regular disruption. Generators and fuel are hard to come by. As a result, universities and research institutes have to plan elaborate schedules for experiments and online teaching, because students in different places don’t always have electricity at the same time.

As Russia continues to bomb the country, the other regular, but less predictable, interruptions are air-raid sirens, which typically give a 30–60 minute warning to move to safety, says Komarov. “Each man and woman has found a way to cope,” he says. “I just keep working, working and working even if things go wrong.”

Komarov, who heads a staff of 60 overseeing 500 undergraduates, says the first month of the war was particularly hard, and dominated by struggles to source even basic reagents. A low point was New Year’s Eve, when Russia bombed his university, shattering more than 1,000 windows that needed to be boarded up before temperatures plunged.



The Taras Shevchenko National University of Kyiv was damaged by bombing.Credit: Taras Shevchenko National University

“Most of my time and that of my colleagues — the deans and directors — is spent on adjusting the working schedule to the power cut-offs and alarms, and organizing logistics, especially for reagents coming from abroad,” he says. In his spare time, Komarov raises funds for the war effort and works on proposals for the construction of a new biomedical science hub when the war is over.



Rebuilding Ukrainian science can’t wait — here’s how to start

“I make a major effort to continue doing science. It’s difficult not only physically but also because of this psychological pressure. To do something creative you need a peaceful time. Even writing papers became quite a challenge.”

Since the war began, his university has lost about one-fifth of its undergraduate and PhD students, says Komarov, who worries about brain drain. (The loss of students is also affecting his university’s funding). “There is a great shortage of scientists here in Ukraine now.” Half of his staff and students are displaced within Ukraine, but are still working; some university staff, he adds, have gone to battlefields, where some have died.

Help for researchers in Ukraine

Last year, Olga Polotska, who heads the National Research Foundation of Ukraine (NRFU), a grant-giving body set up in 2018, was looking forward to giving out her agency’s largest ever funding total, 27 million (US$29 million), double that of the previous year, in project grants. But shortly after the invasion, the government decided to redirect those funds to help fight the war. Now, the NRFU is starting to honour its 2022 commitments. Polotska had initially focused her efforts on drumming up international support for

Ukrainian researchers abroad. “But the time came, in early summer, when I told myself I should focus on those who are remaining,” she says. “We need these people,” she adds. “Scientists are the golden resource.”



Olga Polotska, who heads the National Research Foundation of Ukraine, speaking at an event where Polish government research funder NCN presented support schemes for Ukrainian researchers.Credit: Jan Bielecki for the NCN

A number of international aid schemes now do support scientists in Ukraine. In June 2022, the Swiss National Science Foundation in Bern extended a scheme supporting Ukrainian scientists who have fled to Switzerland, to allow remote funding for projects in Ukraine, between teams from the two countries. And the Dutch Research Council (NWO) is discussing a €250,000 scheme under which Ukrainian researchers can remotely join existing NWO-financed projects. Last December, the University of Cambridge, UK, and the NRFU launched a scheme under which the university will fund up to 15 grants for individual researchers from Ukraine who have been displaced by the war and are still living in Ukraine.

“To my mind it’s perfect,” says Polotska, “to provide grant support to researchers staying in Ukraine from the regions that have been severely affected by the military aggression.”

In January, the Simons Foundation in New York City, which supports research in mathematics and basic science, announced more than $1.2 million of funding for 405 scientists in Ukraine, including doctoral candidates. Each will receive a $100 or $200 monthly stipend with some leaders of research teams receiving larger, lump sums.



Ukraine’s scientists need help to rebuild their research system

“The main point [of focusing grants into the country] is to assure that the scientists who are in Ukraine, can still remain scientists. So that when the war eventually finishes, they have not left academia altogether and not lost contact with the international scientific community,” says Michael Rose, who studies the economics of science and innovation at the Max Planck Institute for Innovation and Competition in Munich, Germany.

Rose is a co-founder of #ScienceForUkraine (SfU), a virtual international network of researchers and students. SfU first focused its efforts on coordinating positions for scientists who had left Ukraine, but it is now also looking to support researchers who have remained in the country. Last September, it established — along with the global initiative Economists for

Ukraine — a programme to award 70 stipends, each worth $5,000, to scientists in Ukraine. The award includes networking and engagement with host institutions, which are mostly US universities.

SfU is also experimenting with a website on which more than 80 Ukrainian research institutes have posted requests to the international community. These range from fire extinguishers to internships, although Rose says responses to the requests have been disappointing so far.

Komarov thinks that international support should focus on helping Ukrainians to keep up their research and teaching within the country. Such programmes “will help us to survive to restore the scientific infrastructure after the war”, he says.

Deepening connections

Polotska says support doesn’t have to be financial. “There have been so many bottom-up initiatives, for example organizing workshops on grant management. One of the key psychological difficulties that people are experiencing is lack of communication, lack of access to the international community.”

Young academics, such as Ivan Brusak, a 28-year-old specialist in geodesy at Lviv Polytechnic National University in Ukraine, say they need equipment and remote international collaborations.

Brusak took months off work when the war began, helping to organize student support for the war effort. His team coordinated a giant weaving operation conducted at more than 100 venues around Lviv, turning tonnes of fabrics and threads into camouflage nets. It also procured and delivered kit such as binoculars and thermal imaging cameras to the military.



Geodesy specialist Ivan Brusak wove camouflage nets for the war effort in 2022.Credit: Katya Moskalyuk/Global Images Ukraine via Getty

Brusak would volunteer from 8 a.m. to 8 p.m., writing scientific papers in the evenings. “Frankly, I still don’t understand how I managed it,” he says. By summer, his time volunteering had dropped and he successfully defended his thesis in December.

His work now takes him about three times as long as it usually would, he says. “I feel constant fatigue from the war around me and the deaths of people I know, which diverts the focus from science.” But, he adds, “my colleagues from Ukraine and I are not going to give up. We are waiting for victory in the war”.

He values a collaboration he has with Polish scientists that began last July. “The understanding that despite the war we continue to cooperate gives motivation.”

Ukrainian scientists abroad

Meanwhile, numerous universities and international collaborations have offered a range of support including spare office space, fellowships and jobs, to help the thousands (or possibly tens of thousands) of scientists who have left Ukraine. Rose says that Poland and Germany have taken in the largest number of scientists, but there are also many in France, Spain and Italy and smaller numbers in the United Kingdom and the United States. One of the largest support schemes was the European Union's 25 million MSCA4Ukraine, which offered temporary placements for doctoral and postdoc students under its Marie Skłodowska-Curie Actions programme.

Rose says that SfU alone has helped about 1,000 scientists find positions outside Ukraine, with about 10 new opportunities still emerging each week.



Scientists in Ukraine have long fought for scientific freedom

Last December, Polish and Ukrainian scientists working with SfU and other institutions released a survey of more than 600 Ukrainian scientists who left their country after the invasion (see go.nature.com/3k3zdpy). The organizers reached as many respondents as they could with the help of networks such as SfU, but say that they probably missed scientists who are not linked to these groups. Still, the results suggest that those outside Ukraine are mostly women (largely because it's harder for men to leave: those aged 18–60 are not supposed to leave the country, although special dispensation can be

granted, such as to single parents). The majority are senior researchers, and many have children with them.

Their relationship with their home institutions in Ukraine is varied.

"We have seen that researchers don't want to abandon their Ukrainian employment," says Iryna Degtyarova, a co-author of the survey report. She works at the SGH Warsaw School of Economics, Poland, in a role created since the war began, building links between the school and Ukrainian universities. "They are not typical refugees who have burned their bridges. It makes a huge difference that the threat is not from within their country: they have a completely different mindset."

The survey found that around one-third were continuing their work in Ukraine remotely. But 30% reported that there were tensions with some home institutions, for example because the institution had told absent teaching staff to return or lose their jobs. And although 42% said they would return to Ukraine if the war ended within a few months, equal proportions of the remainder said they would not return or didn't know what they would do.

The important thing, says Degtyarova, is to keep Ukrainians abroad connected to their work back home.

For some, ties have already loosened. Olena Prysiazhna, a 35-year-old specialist in plasma physics who was at the National University of Kyiv, has left Ukrainian science — at least for now — after her Kyiv contract lapsed when she couldn't get back in time to renew it. She had fled with her mother, younger sister (a physics postgraduate student) and their dog, after a missile landed in their neighbour's back garden. When *Nature* spoke to Prysiazhna last April (see *Nature* **605**, 414–416; 2022), the family was in the Netherlands recovering from the ordeal, and both sisters were looking for scientific work.



Plasma physicist Olena Prysiazhna left Ukraine and is now working in Japan.Credit: Ilvy Njiokiktjien/VII Photo

Now, through a contact she had previously made at a conference, Prysiazhna has a part-time job at Sojo University in Kumamoto, Japan, in a different field of plasma physics. “It’s hard to be alone,” she says, saying that she checks Facebook every morning to make sure her old friends and colleagues in Ukraine are still alive. “But war has changed me. You become more brave and resilient,” she says. “If someone had told me a year ago I would go to Japan on my own I would not have believed them. But I had no fear.”

Prysiazhna assumes she will eventually return to Ukraine, although her family’s home is damaged, as her mother found on a brief return visit. Prysiazhna’s mother and sister, and the dog, now share a room in a converted holiday camp in the Netherlands. Her sister, who had almost finished her PhD when she left Ukraine, is trying to find a way to defend her thesis and is working for a logistics company.

Komarov’s worry is that scientists in Prysiazhna’s position — young and with few ties — might never return. But Rose, who studies phenomena such

as brain drain, disagrees.

“Irrespective of whether people are refugees, irrespective of whether they return or not,” he says, “these fears of brain drain are not substantiated. Most people return and then they bring new ideas back. But even when they don’t return, they still pass on knowledge and insight to people who are still there. Even this way there is knowledge sharing.”

Whether the aid is inside or outside the country, Rose, Bondar and others are convinced that, out of the horrors of the Russian invasion, there could emerge a web of connections with Western science that will help Ukraine — if it wins the war — to rebuild and reform its science, and build a future more closely allied with Europe.

Komarov and Minakova both want to keep Ukrainian science strong now, so that the research community can make rapid gains when the war ends. Minakova sees her research as crucial to this effort. Her team and Bondar’s have just won a grant from the US Embassy in Kyiv to fund the education of master’s students in the science, technology, policy and business of solar power.

“The war will stop and then we will need to be good specialists in alternative energy,” she says. “We are going to rebuild our country and our campuses and I hope the world will help us.”

Nature **614**, 608-612 (2023)

doi: https://doi.org/10.1038/d41586-023-00508-0

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00508-0

- NEWS FEATURE
- 22 February 2023

Your brain could be controlling how sick you get — and how you recover

Scientists are deciphering how the brain choreographs immune responses, hoping to find treatments for a range of diseases.

- Diana Kwon



Illustration by Dalbert B. Vilarino

Hundreds of scientists around the world are looking for ways to treat heart attacks. But few started where Hedva Haykin has: in the brain.

Haykin, a doctoral student at the Technion — Israel Institute of Technology in Haifa, wants to know whether stimulating a region of the brain involved in positive emotion and motivation can influence how the heart heals.

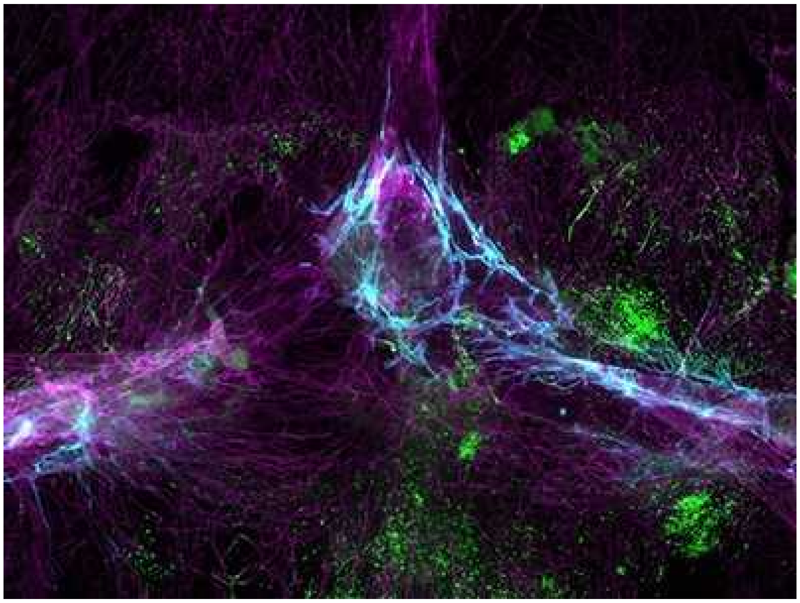
Late last year, in a small, windowless microscope room, she pulled out slides from a thin black box, one by one. On them were slices of hearts, no bigger than pumpkin seeds, from mice that had experienced heart attacks. Under a microscope, some of the samples were clearly marred by scars left in the aftermath of the infarction. Others showed mere speckles of damage visible among streaks of healthy, red-stained cells.

The difference in the hearts' appearance originated in the brain, Haykin explains. The healthier-looking samples came from mice that had received stimulation of a brain area involved in positive emotion and motivation. Those marked with scars were from unstimulated mice.

"In the beginning we were sure that it was too good to be true," Haykin says. It was only after repeating the experiment several times, she adds, that she was able to accept that the effect she was seeing was real.

Haykin, alongside her supervisors at the Technion — Asya Rolls, a neuroimmunologist, and Lior Gepstein, a cardiologist — are trying to work out exactly how this happens. On the basis of their experiments so far, which have not yet been published, activation of this brain reward centre — called the ventral tegmental area (VTA) — seems to trigger immune changes that contribute to the reduction of scar tissue.

This study has its roots in decades of research pointing to the contribution of a person's psychological state to their heart health[1]. In a well-known condition known as 'broken-heart syndrome', an extremely stressful event can generate the symptoms of a heart attack — and can, in rare cases, be fatal. Conversely, studies have suggested that a positive mindset can lead to better outcomes in those with cardiovascular disease. But the mechanisms behind these links remain elusive.

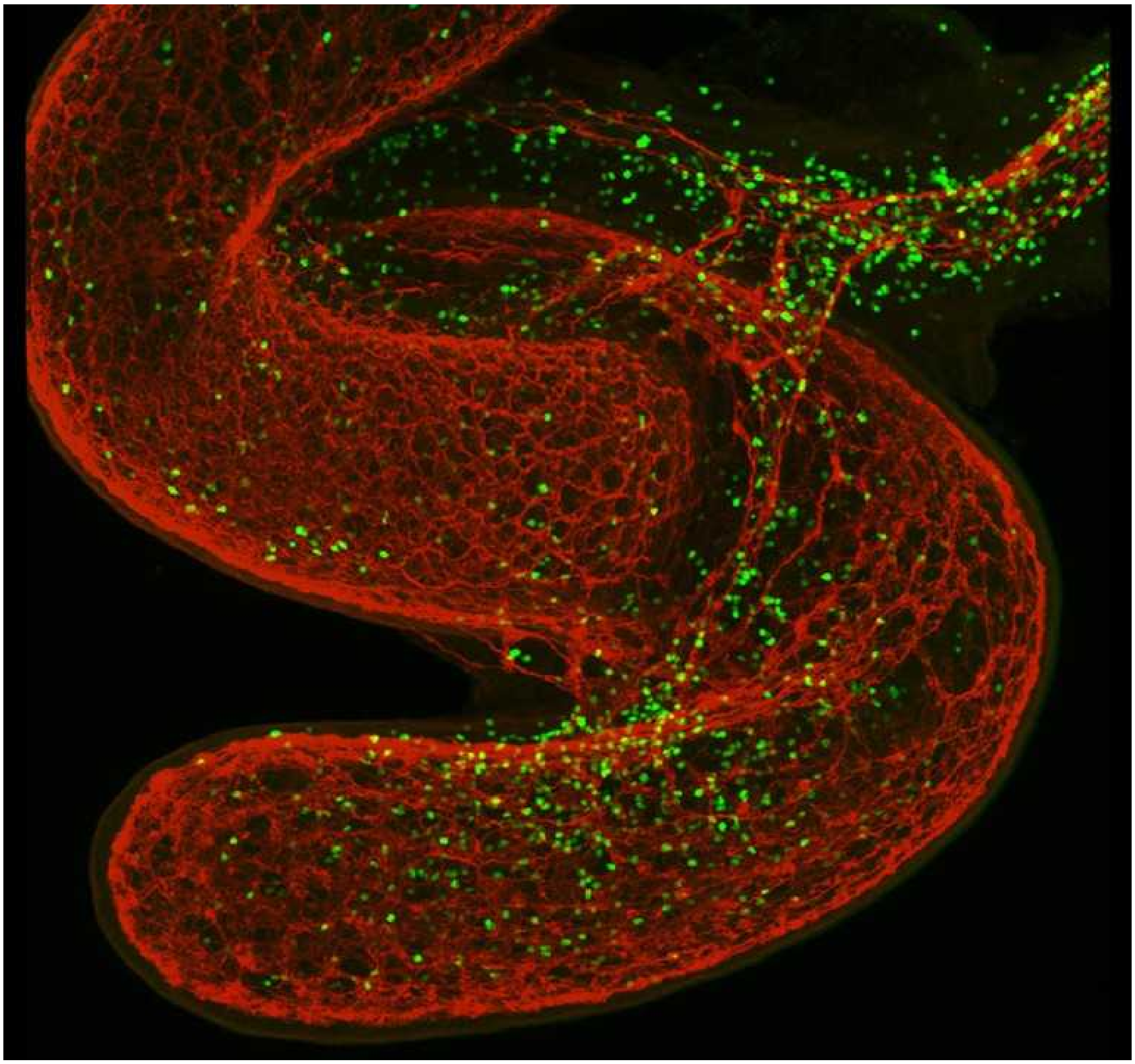


[Guardians of the brain: how a special immune system protects our grey matter]

Rolls is used to being surprised by the results in her laboratory, where the main focus is on how the brain directs the immune response, and how this connection influences health and disease. Although Rolls can barely contain her excitement as she discusses her group's eclectic mix of ongoing studies, she's also cautious. Because of the often-unexpected nature of her team's discoveries, she never lets herself believe an experiment's results until they have been repeated multiple times — a policy that Haykin and others in her group have adopted. "You need to convince yourself all the time with this stuff," Rolls says.

For Rolls, the implications of this work are broad. She wants to provide an explanation for a phenomenon that many clinicians and researchers are aware of: mental states can have a profound impact on how ill we get — and how well we recover. In Rolls's view, working out how this happens could enable physicians to tap into the power of the mind over the body. Understanding this could help to boost the placebo effect, destroy cancers, enhance responses to vaccination and even re-evaluate illnesses that, for centuries, have been dismissed as being psychologically driven, she says. "I think we're ready to say that psychosomatic [conditions] can be treated differently."

She is part of a growing group of scientists who are mapping out the brain's control over the body's immune responses. There are multiple lines of communication between the nervous and the immune systems — from small local circuits in organs such as the skin, to longer-range routes beginning in the brain — with roles in a wide range of diseases, from autoimmunity to cancer. This field "has really exploded over the last several years", says Filip Swirski, an immunologist at the Icahn School of Medicine at Mount Sinai in New York City.



Neuronal cells (red) in the gut interface with cells of the immune system (green).Credit: Lara Santos/Veiga-Fernandes Lab

Some parts of the system — such as the vagus nerve, a huge highway of nerve fibres that connects the body to the brain — have inspired treatments for several autoimmune diseases that are currently being tested in clinical trials. Other studies, investigating how to recruit the brain itself — which some think could provide powerful therapies — are still nascent. Rolls, for one, has just begun examining whether the pathways her team has found in mice are also present in humans. And she has launched a start-up company to try to develop treatments based on her findings.

Although these developments are encouraging to researchers, much is still a mystery. “We often have a black box between the brain and the effect we see in the periphery,” says Henrique Veiga-Fernandes, a neuroimmunologist at the Champalimaud Centre for the Unknown in Lisbon. “If we want to use it in the therapeutic context, we actually need to understand the mechanism.”

A tale of two systems

For more than a century, scientists have been finding hints of a close-knit relationship between the nervous and the immune systems. In the late nineteenth and early twentieth centuries, for example, scientists demonstrated that cutting nerves to the skin could curb some hallmarks of inflammation[2].

It wasn’t until the late 1990s that researchers in this field began drawing connections to the body’s master conductor, the brain. Neurosurgeon Kevin Tracey, then at North Shore University Hospital in Manhasset, New York, and his colleagues found something unexpected while investigating whether an experimental anti-inflammatory drug could help to tame brain inflammation caused by stroke.



How gut microbes could drive brain disorders

When delivered into the brains of rodents that had experienced strokes, the drug had the expected effect: it reduced neuroinflammation. As a control, the team injected the drug into the brains of animals that had inflammation throughout their bodies, thinking the drug would work exclusively in the brain. To their surprise, it also worked in the body. “This was a real head-scratcher,” says Tracey, now president and chief executive of the Feinstein Institutes for Medical Research in Manhasset.

After months of trying to determine the path of the drug from brain to body, the researchers decided to cut the vagus nerve, a bundle of some 100,000 nerve fibres that runs from the brain to the heart, lungs, gastrointestinal tract and other major organs. With the vagus nerve snipped, the anti-inflammatory effect of the brain-administered drug disappeared[3].

Inspired by this discovery, Tracey’s group and others have continued to explore other ways in which the vagus nerve — and the rest of the nervous system — directs immune responses. A driving force for these developments, says Swirski, has been the advent of scientific tools that enable scientists to begin to chart the interactions between the nervous and the immune systems in an unprecedented way.

Some researchers are focusing on particular body systems. For instance, a team led by Andreas Habenicht, a cardiologist at LMU Munich, Germany, reported last year that the interaction between immune cells and nerves in the outermost layer of artery walls modulated the progression of atherosclerosis, an inflammatory disease in which vessels become clogged with cholesterol and other substances[4].

Meanwhile, Veiga-Fernandes and his group have documented clusters of neuronal and immune cells in various tissues and discovered how they work together to sense damage and mobilize immune reactions. His team is now looking at how these little switchboards can be controlled by the brain[5].



Neuroscientists Catherine Dulac (right) and Jessica Osterhout with images of neurons in the hypothalamus that control symptoms of sickness, such as fever and loss of appetite.Credit: Kris Snibbe/Harvard Staff Photographer

The brain itself is also beginning to give up its secrets. Neuroscientist Catherine Dulac and her team at Harvard University in Cambridge, Massachusetts, have pinpointed neurons in an area called the hypothalamus that control symptoms including fever, warmth-seeking and loss of appetite

in response to infection[6]. “Most people probably assume that when you feel sick, it’s because the bacteria or viruses are messing up your body,” she says. But her team demonstrated that activating these neurons could generate symptoms of sickness even in the absence of a pathogen. An open question, Dulac adds, is whether these hypothalamic neurons can be activated by triggers other than pathogens, such as chronic inflammation.

Just above the hypothalamus sits a region called the insula, which is involved in processing emotion and bodily sensations. In a 2021 study, one of Rolls’s doctoral students, Tamar Koren, found that neurons in the insula store memories of past bouts of gut inflammation — and that stimulating those brain cells reactivated the immune response[7].

Rolls, Koren and their colleagues suspect that such a reaction might prime the body to fight potential threats. But these reactions could also backfire and start up in the absence of the original trigger. This could be the case for certain conditions, such as irritable bowel syndrome, that can be exacerbated by negative psychological states.

Mind over matter

Many scientists hope to pin down how such mental states influence immune responses.

Rolls and Fahed Hakim, a paediatrician and director of the Nazareth Hospital EMMS in Israel, were inspired to investigate this question after coming across a 1989 study[8] reporting that, among women with breast cancer, those who underwent supportive group therapy and self-hypnosis in addition to routine cancer care survived longer than those who received only the latter. Several other studies have documented a similar link between survival and the mental states of people with cancer.

To test the link, Rolls, Hakim and their team zoomed in on the VTA — the same region they targeted in the heart-attack study and in a previous experiment looking at bacterial infection. This time they focused on mice with lung and skin tumours. Activating neurons in the VTA noticeably shrank the cancers[9]. It turned out that VTA activation subdued cells in the

bone marrow that would usually repress immune activity, freeing the immune system to fight the cancer.



Neuroscientist Talma Hendler readies a participant for a brain scan. The results will inform a study of whether learning to control brain activity can improve the immune response to a vaccine.Credit: Sourasky Medical Center Lior Zur

Clinicians have known about the effect of positive thinking on disease progression for a long time, Hakim says. But this evidence has been largely anecdotal or correlational, so being able to identify a pathway through which such an effect occurs — and manipulate it experimentally in animals — makes it much more real, he says.

Negative mental states can also influence the body's immune response. In a study published last year, Swirski and his team identified specific brain circuits that mobilize immune cells in the bodies of mice during acute stress[10]. The researchers found two pathways, one originating in the motor cortex that directed immune cells to the site of injury or infection, and another beginning in the hypothalamus — a key responder in times of stress

— that reduced the number of immune cells circulating in the blood. The group is now investigating the role of stress-mediated circuits in chronic inflammatory diseases.

Neuroscientist Jeremy Borniger at Cold Spring Harbor Laboratory in New York and his colleagues have also found that activating neurons in the mouse hypothalamus can generate an immune response[11] — and are now examining how manipulating these cells can alter the growth of tumours.

Some groups are hoping to replicate their findings in humans. Swirski's team, for instance, plans to use tools such as virtual reality to manipulate people's stress levels and see how that changes the immune response.

Koren and Rolls are working with Talma Hendler, a neuroscientist and psychiatrist at Tel Aviv University in Israel, to see whether boosting the reward system in people's brains before they receive a vaccine can improve their immune response. Rather than stimulating the brain directly, they are using a method called neurofeedback, in which individuals learn to observe and control their own brain activity; this is measured using methods such as functional magnetic resonance imaging.

The road to the clinic

Over the years, Rolls would chat with her good friend Tehila Ben-Moshe about her research. Ben-Moshe is the chief executive of Biond Biologics, an Israel-based biopharmaceutical company that focuses on using immune cells to target cancer. During one such discussion last year, Ben-Moshe realized that Rolls's brain-stimulation experiments were acting on some of the same immune cells that her company was trying to target, and immediately saw the therapeutic potential. "When I saw Asya's data, I couldn't believe what I saw," says Ben-Moshe. "The question then became — how can I translate what she's doing with mice into patients?" The two are working on launching a company.



Nature Reviews Immunology: Series on neuroimmunology

Ben-Moshe and Rolls hope to harness existing brain-stimulation technologies, such as transcranial magnetic stimulation, which uses magnetic pulses to alter brain activity, or focused ultrasound, which uses sound waves, to modulate the immune systems of people with cancer, autoimmune diseases or other conditions. As a first step, their team has been reaching out to companies that have developed such technologies. Before starting clinical trials, Ben-Moshe and Rolls want to examine blood samples from trials already performed with these techniques, to see whether there are signs of immune-system alterations before and after treatment.

Potential therapies targeting the vagus nerve are nearer the clinic. A company co-founded by Tracey — SetPoint Medical in Valencia, California — is testing pill-sized vagus-nerve stimulators, implanted in the vagus nerve in the neck, in autoimmune diseases including Crohn's disease, multiple sclerosis and rheumatoid arthritis. The rheumatoid-arthritis trial is farthest along — the team has shown in a small trial in Europe that its device can reduce disease severity[12]. The technique is currently undergoing a randomized, sham-controlled trial (in which the control group will receive an implant but no active stimulation) in 250 patients in various centres across the United States.

Rolls's hope is that this work will ultimately help physicians to understand, and act on, the mind–body connections that they see in their practices. The need is clear: when Rolls put out a call to speak to psychologists from the hospital where her lab is based, the meeting room was packed. People from departments ranging from dermatology to oncology were eager to share their stories. Many clinicians pass people with seemingly psychosomatic issues on to psychologists, saying there is nothing physically wrong, said one attendee. This can be distressing for the person seeking treatment. Even being able to simply tell people that there is a brain–immune connection that is responsible for their symptoms can make an enormous difference.

It's time that both researchers and clinicians take the link between psychology and physiology seriously, says Rolls. "You can call something psychosomatic, but in the end, it's somatic. How long can we ignore what is there?"

Nature **614**, 613-615 (2023)

doi: https://doi.org/10.1038/d41586-023-00509-z

References

1. Everson-Rose, S. A. & Lewis, T. T. *Annu. Rev. Public Health* **26**, 469–500 (2005).
2. Sousa-Valente, J. & Brain, S. *Semin. Immunopathol.* **40**, 229–236 (2018).
3. Borovikova, L. *et al. Nature* **405**, 458–462 (2000).
4. Mohanta, S. K. *et al. Nature* **605**, 152–159 (2022).
5. Cardoso, F. *et al. Nature* **597**, 410–414 (2021).
6. Osterhout, J. A. *et al. Nature* **606**, 937–944 (2022).
7. Koren, T. *et al. Cell* **184**, 5902–5915 (2021).

8. Spiegel, D. *et al. Lancet* **334**, 888–891 (1989).

9. Ben-Shaanan, T. L. *et al. Nature Commun.* **9**, 2723 (2018).

10. Poller, W. C. *et al. Nature* **607**, 578–584 (2022).

11. Li, S.-B. *et al. Sci. Adv.* **6**, eabc2590 (2020).

12. Koopman, F. A. *et al. Proc. Natl Acad. Sci. USA* **113**, 8284–8289 (2016).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00509-z

Books & Arts

- **Hauling icebergs to Africa: could a bizarre plan to get drinking water actually work?** [20 Feb 2023]
  • Transporting water trapped in icebergs to drought-plagued regions is pooh-poohed by scientists — but some see it as a huge opportunity.
- **The failure of ‘democratized’ health care, and parasites on parade: Books in brief** [25 Jan 2023]
  • Andrew Robinson reviews five of the best science picks.

- BOOK REVIEW
- 20 February 2023

Hauling icebergs to Africa: could a bizarre plan to get drinking water actually work?

Transporting water trapped in icebergs to drought-plagued regions is pooh-poohed by scientists — but some see it as a huge opportunity.

- Josie Glausiusz



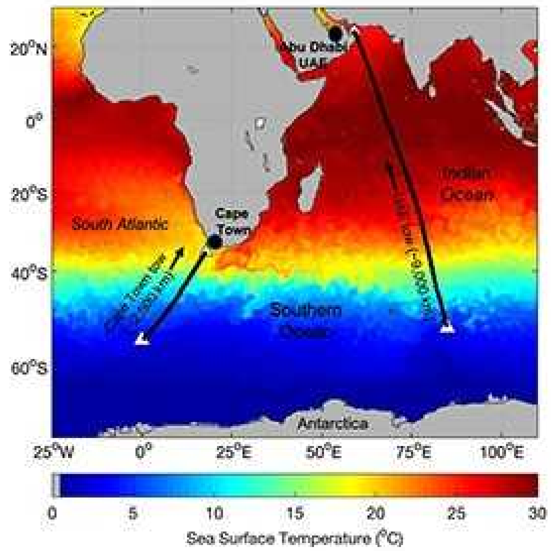
Two-thirds of global fresh water is locked away in ice. Credit: Mario Tama/Getty

Chasing Icebergs: How Frozen Freshwater Can Save the Planet
Matthew H. Birkhold Viking (2023)

The flavour of chilled Svalbarði-brand water, melted from an iceberg just 1,000 kilometres from the North Pole, is described by the company as "like catching snowflakes on the tongue". Bottled in Longyearbyen, a tiny metropolis in Norway's Svalbard archipelago, Svalbarði water is airlifted to luxury locales in London, Sydney, Florida and Macau. "Taste the Arctic to Save the Arctic," its website croons, promoting the supposed carbon neutrality of the water, which sells online for €99.95 (US$107) for a 750-millilitre bottle.

That price is far out of reach for most of the world's people, including the one in four who lack safe drinking water. In *Chasing Icebergs*, Matthew Birkhold, a scholar of law, culture and the humanities, considers whether it's possible to slake the world's thirst with the two-thirds of global fresh water that is locked away in ice caps and glaciers — "stuck at the poles in gigantic fortresses of ice", in his words.

Some are already at it — and not just epicureans quaffing Svalbarði. In Newfoundland, Canada, Birkhold interviews "iceberg cowboy" Ed Kean, who wrangles bergs from the frigid sea, selling the water to cosmetics companies and breweries. In Qaanaaq, Greenland's northernmost town, Birkhold notes, the public water supply includes filtered and treated iceberg melt.



[Towing icebergs to arid regions to reduce water scarcity]

There’s more where that came from. By one estimate, some 2,300 cubic kilometres of ice breaks off from Antarctica every year. More than 100,000 Arctic and Antarctic icebergs melt into the ocean annually, [according to a 2022 United Nations report]. A relatively small, 113-million-tonne iceberg, says Birkhold, could be towed from Antarctica to Cape Town, South Africa, to supply 20% of the city’s water needs for a year. What’s not to like?

Quite a lot, perhaps. “To write this book, I have talked to dozens of scientists,” Birkhold writes. “They are uniformly dubious.” Palaeoclimatologist Ellen Mosley-Thompson has led nine expeditions to Antarctica and six to Greenland to extract ice cores. “Make sure you write that I am skeptical of iceberg towing,” she instructs him.

Less sceptical is master mariner Nick Sloane, the brains behind the Cape Town plan. Using satellite data to find the best berg, his “team of glaciologists, engineers and oceanographers” plans to catch it in a giant net and pull it by tugboat into the mighty Antarctic Circumpolar Current, and thence into the north-flowing Benguela Current towards South Africa.



Icebergs are sometimes towed to prevent damage to oil platforms.Credit: Greg Locke/Reuters

Sloane has estimated the cost at a cool $100 million, plus an extra $50 million or so to melt the ice and funnel the fresh water to land, if the iceberg hasn't melted into the sea or fallen apart en route. In various interviews, Cape Town officials have expressed themselves as less than enthused.

Sloane's plans have yet to materialize. Meanwhile, POLEWATER, a Berlin-based company, has been working for almost a decade on a similar plan — to haul frozen fresh water to the western coast of Africa and the Caribbean, where the company intends to give it away to those in dire need. It, too, will use satellites to locate suitable bergs, but once it has relocated them, it plans to pump meltwater from pools on top into easily transportable gigantic bags.

Then there's the UAE Iceberg Project, the dream of Emirati inventor Abdulla Alshehhi to import an Antarctic iceberg to the Fujairah coast of the United Arab Emirates. An animated promotion features penguins and polar bears — species hailing from opposite poles — posing dolefully on the

iceberg. Alshehhi has said that “it will be cheaper to bring in these icebergs” than to desalinate seawater — a common technique in the Middle East.



World’s largest ice sheet threatened by warm water surge

Desalination provides at least 35 trillion litres of drinking water globally every year. Birkhold notes that it is prohibitively expensive in many places. It also relies on fossil fuels for energy, and pollutes the ocean with excess salt. But he offers limited information on other, perhaps more effective, alternatives to iceberg towing, such as recycling municipal wastewater or tapping brackish water for crop irrigation. He offers no data on more esoteric sources of water, such as fog harvesting, used by remote communities in Chile, Morocco and South Africa. Nor does he address initiatives to reduce water waste or increase efficiency of use.

Birkhold speculates that if iceberg harvesting succeeds, it might not remain the province of quixotic entrepreneurs — bigger, water-hungry enterprises might muscle in with their “deep coffers and profit-driven approach”. The race is largely unregulated: few national laws address iceberg use, and no international agreements clarify who can capture and sell these freshwater resources. If they are to be exploited fairly, the author concludes, “we need to decide who gets to use icebergs — and how and how many — in a way that is just and equitable”.

Birkhold is engagingly honest about potential pitfalls. But if the enterprise could succeed, enormous quantities of fresh water that would otherwise melt into the ocean could be delivered to parched regions. Kudos to the author for diving in — whether or not it is ever realized.

Nature **614**, 617-618 (2023)

doi: https://doi.org/10.1038/d41586-023-00510-6

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00510-6

| Section menu | Main menu |

- BOOK REVIEW
- 25 January 2023

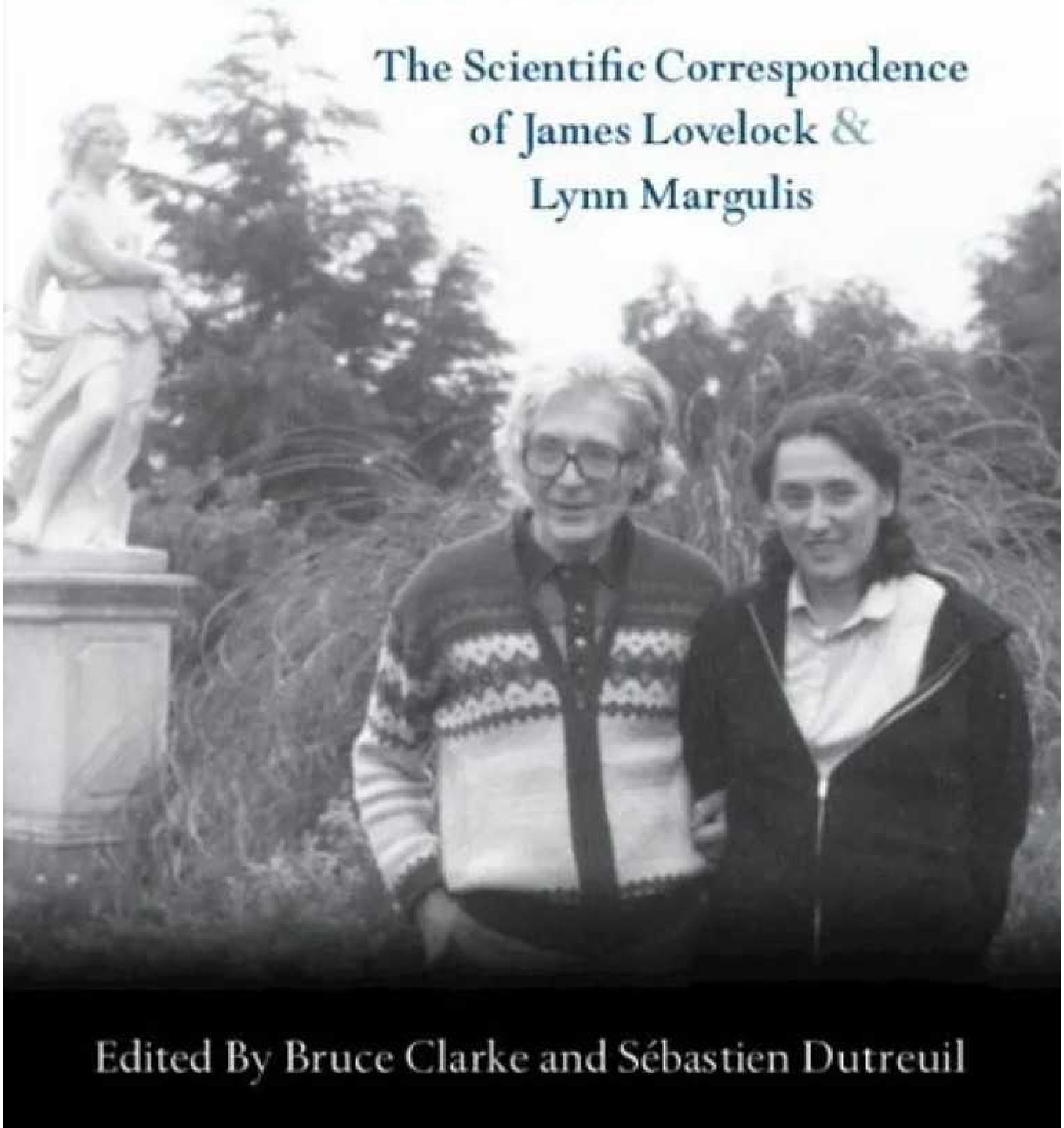
The failure of ‘democratized’ health care, and parasites on parade: Books in brief

Andrew Robinson reviews five of the best science picks.

- Andrew Robinson

Writing Gaia

The Scientific Correspondence of James Lovelock & Lynn Margulis

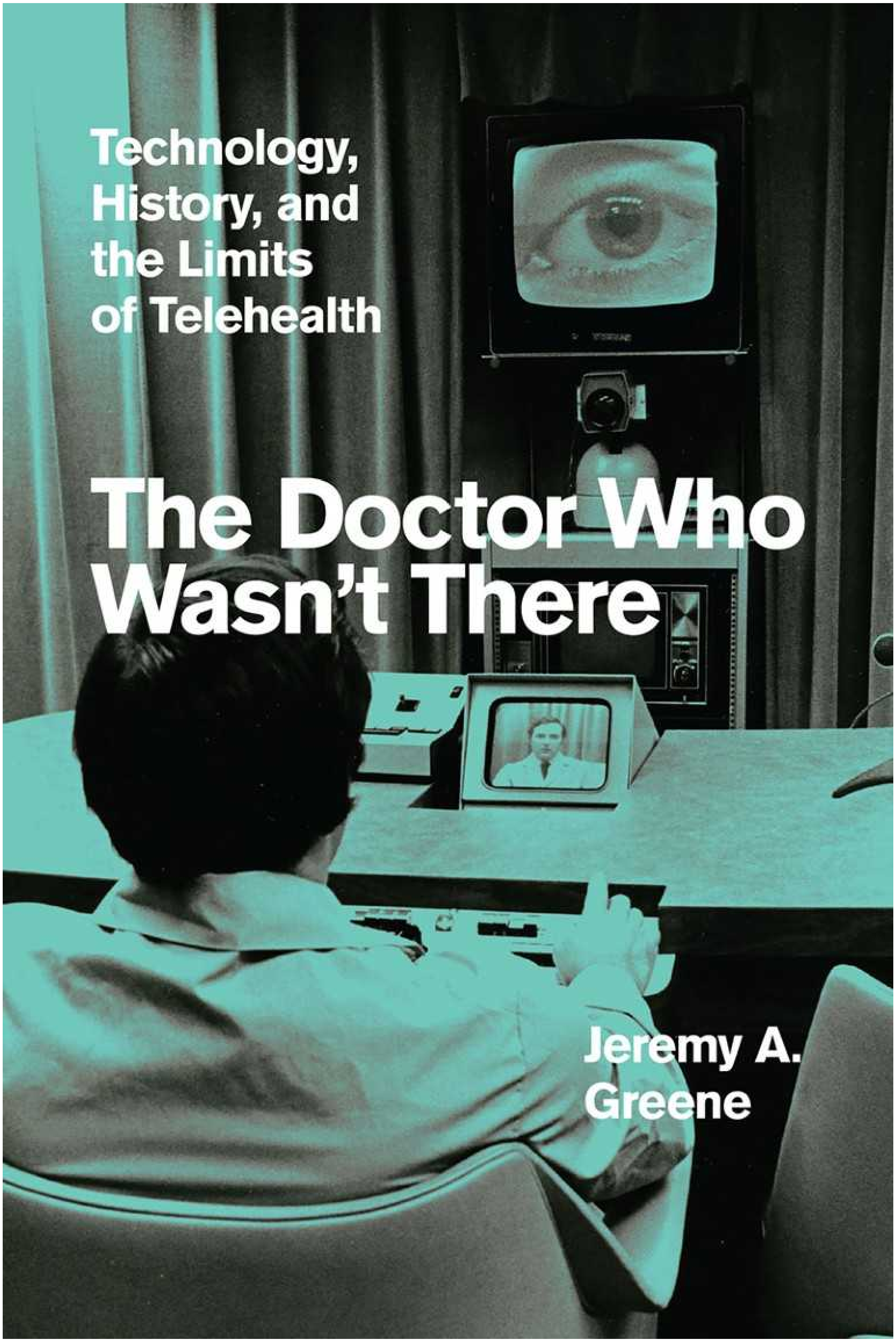


Edited By Bruce Clarke and Sébastien Dutreuil

Writing Gaia

Eds Bruce Clarke & Sébastien Dutreuil *Cambridge Univ. Press* (2022)

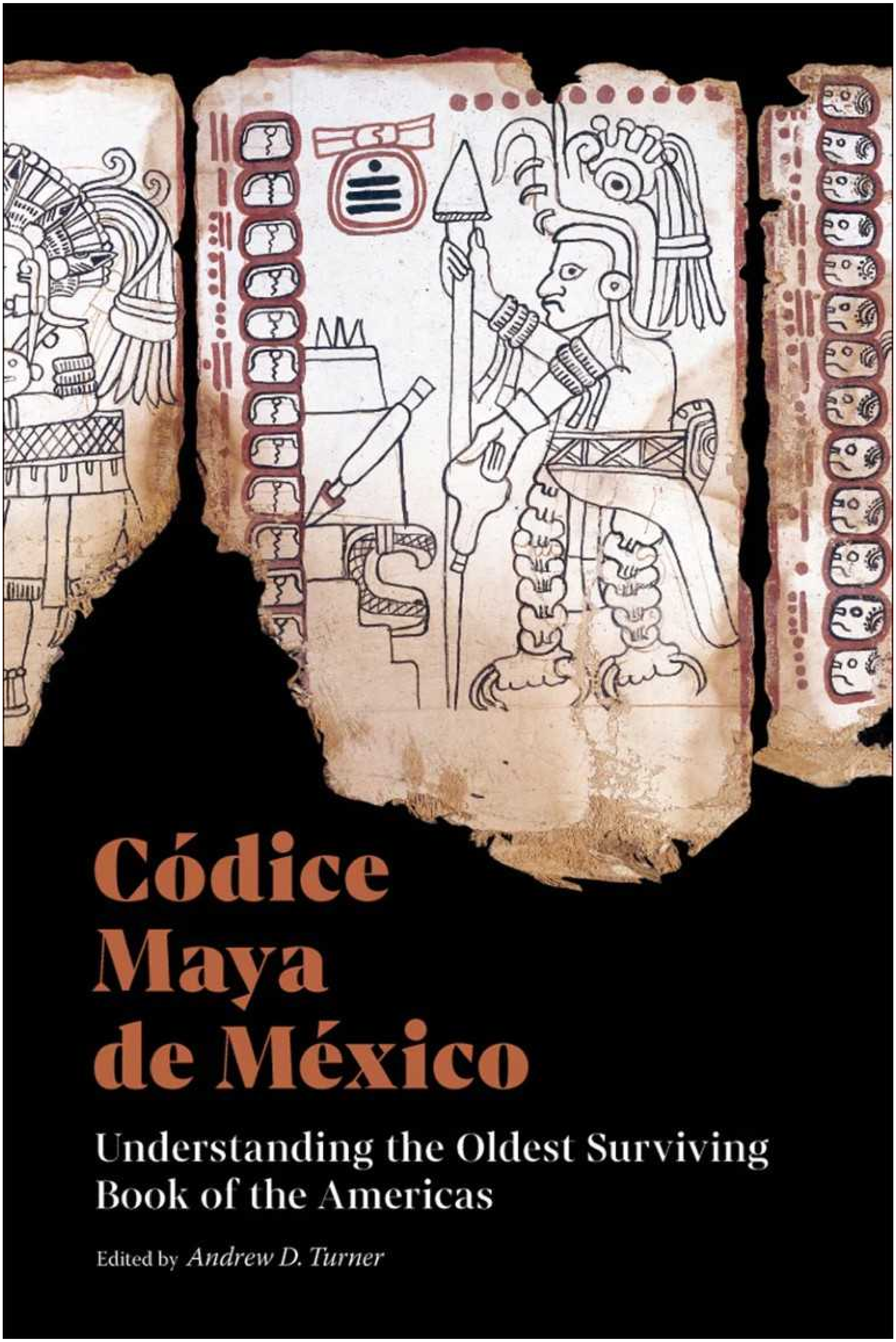
In 1966, independent scientist James Lovelock proposed that Earth's climate and the chemistry of its surface, air and sea "have evolved with life to provide optimum conditions for its survival". From 1972, he and evolutionary biologist Lynn Margulis developed this concept into the Gaia hypothesis. Their trans-Atlantic correspondence appears in this enlightening book, edited by historians Bruce Clarke and Sébastien Dutreuil. At the time, the idea had little backing from Earth and life scientists; now, notables in the field contribute essays of support.


Technology, History, and the Limits of Telehealth
The Doctor Who Wasn't There
Jeremy A. Greene

The Doctor Who Wasn't There

Jeremy A. Greene *Univ. Chicago Press* (2022)

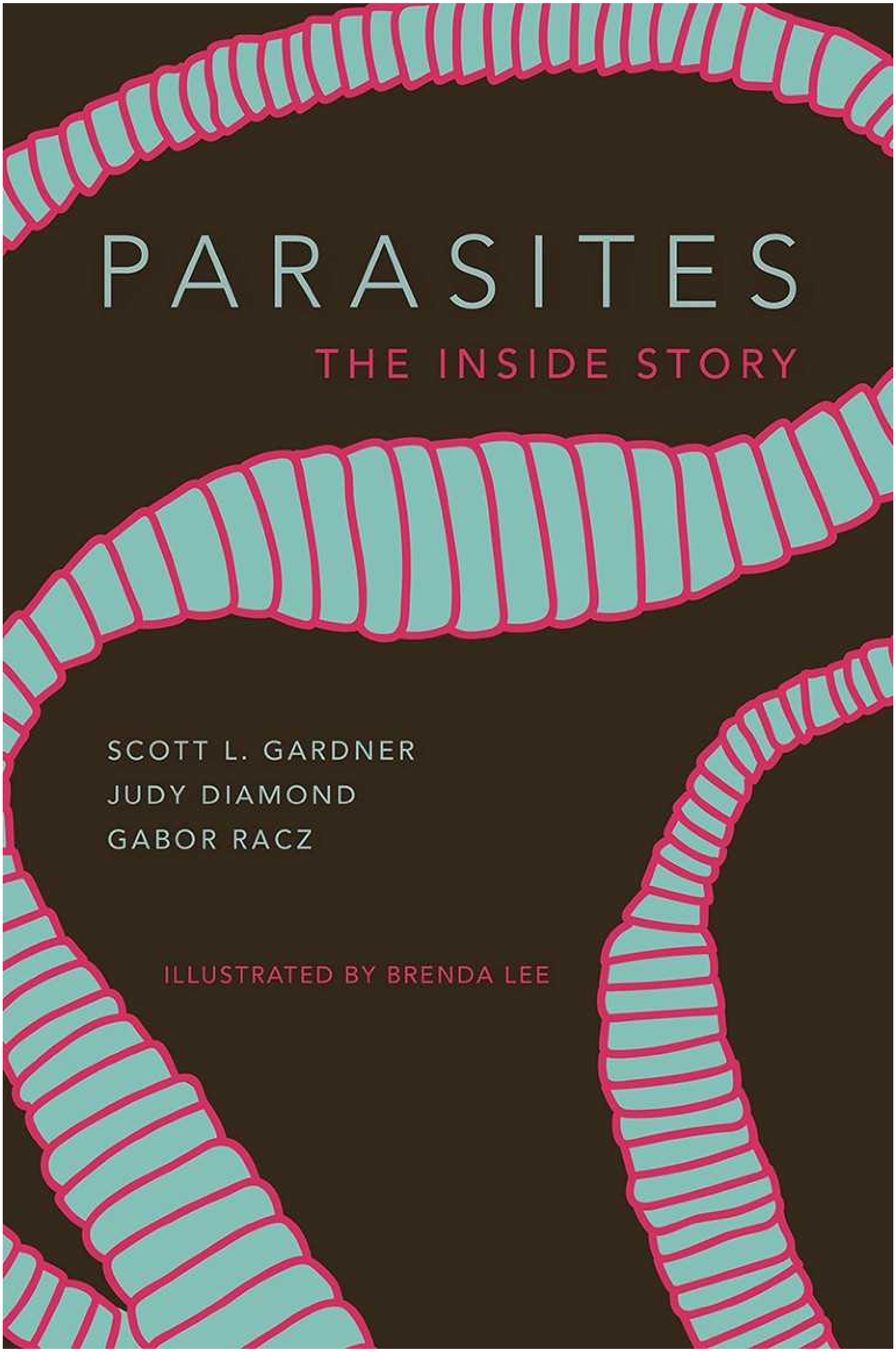
Early in the COVID-19 pandemic, physician Jeremy Greene converted his clinic to telemedicine, with mixed results. He became curious about the medical history of communications technology such as the telephone, invented in 1876. The 1910 'telephonic stethoscope' was hailed for monitoring hearts, lungs and abdomens. Yet it was no use to the numerous patients who lacked a telephone even decades later. This vibrant, highly readable but US-focused history describes how many other technologies tried "to democratize access to healthcare".


Códice
Maya
de México
Understanding the Oldest Surviving
Book of the Americas
Edited by Andrew D. Turner

Códice Maya de México

Ed. Andrew D. Turner *Getty Research Inst.* (2022)

Only four Maya books survived the Spanish conquest of Mesoamerica. Three are in Europe and proved key to deciphering Maya glyphs. The fourth, discovered in the 1960s, remains in Mexico. Its authenticity was debated until 2018, when scrutiny of its colours, notably ‘Maya blue’, established it as the oldest surviving book of the Americas, from around 1100. Archaeologist Andrew Turner and contributors provide a fascinating analysis of this science and the book’s supernatural content — predictions based on Venus’s movements — with a vivid facsimile.


PARASITES
THE INSIDE STORY
SCOTT L. GARDNER
JUDY DIAMOND
GABOR RACZ
ILLUSTRATED BY BRENDA LEE

Parasites

Scott L. Gardner *et al*. *Princeton Univ. Press* (2022)

Parasite originally meant ‘next to food’, but later signified sitting next to a host to get free food. Now it describes a long-term relationship between species, in which the parasite benefits and the host is harmed, although in practice the latter ranges from “deadly effects” to “some benefits”, note parasitologists Scott Gardner and Gabor Rácz, and parrot specialist Judy Diamond. Their world-ranging study, illustrated by Brenda Lee, focuses on three abundant types of parasite — nematodes, flatworms and thorny-headed worms.

The Pandemic Divide

HOW COVID INCREASED INEQUALITY IN AMERICA



Gwendolyn L. Wright,
Lucas Hubbard, and
William A. Darity Jr.,
EDITORS

The Pandemic Divide

Ed. Gwendolyn L. Wright *et al*. *Duke Univ. Press* (2022)

US Black and Latinx populations experience illness and COVID-19 deaths at a higher rate and earlier age than the national average. "COVID-19 pulled back the curtain on the extent of racialized inequality in the United States," notes the foreword to this disturbing but proactive collection of articles by 27 US academics from many fields. So did George Floyd's murder in 2020. Both events highlight inequality arising from discrimination in health care, policing, education, employment, housing and credit markets.

Nature **614**, 618 (2023)

doi: https://doi.org/10.1038/d41586-023-00211-0

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00211-0

| Section menu | Main menu |

Opinion

- **AI weapons: Russia’s war in Ukraine shows why the world must enact a ban** [21 Feb 2023]
  • Conflict pressures are pushing the world closer to autonomous weapons that can kill without human control. Researchers and the international community must join forces to prohibit them.

- COMMENT
- 21 February 2023

AI weapons: Russia’s war in Ukraine shows why the world must enact a ban

Conflict pressures are pushing the world closer to autonomous weapons that can kill without human control. Researchers and the international community must join forces to prohibit them.

- Stuart Russell



A Russian-launched, Iranian Shahed-136 loitering missile flies over Kyiv in October 2022.Credit: Yasuyoshi Chiba/AFP via Getty

One year since Russia's invasion, an arms race in artificial-intelligence (AI) weaponry is being played out on Ukrainian soil. Western audiences cheer when plucky Ukrainian forces use modified commercial quadcopters to drop grenades on Russian soldiers. They boo when brutal Russian forces send swarms of cheap Iranian cruise missiles to destroy hospitals, power plants and apartment blocks. But this simple 'us versus them' narrative obscures a disturbing trend — weapons are becoming ever smarter.

Soon, fully autonomous lethal weapon systems could become commonplace in conflict. Some are already on the market. Mercifully, few have actually been used in warfare, and none has been used in Ukraine, at the time of writing. Yet evolving events are a cause for concern.

The inevitable logic of using electronic countermeasures against remotely operated weapons is driving both sides towards increasing the level of autonomy of those weapons. That is pushing us ever closer to a dangerous world where lethal autonomous weapon systems are cheap and widely available tools for inflicting mass casualties — weapons of mass destruction found in every arms supermarket, for sale to any dictator, warlord or terrorist.

Although it is difficult to discuss banning weapons that might help the Ukrainian cause, it is now urgent that world governments do so and limit the use of AI in war. No one wants this bleak future of robotic threats.

As a start, governments need to begin serious negotiations on a treaty to ban anti-personnel autonomous weapons, at the very least. Professional societies in AI and robotics should develop and enforce codes of conduct outlawing work on lethal autonomous weapons. And people the world over should understand that allowing algorithms to decide to kill humans is a terrible idea.

Pressures leading to full autonomy

What exactly are ‘lethal autonomous weapons systems’? According to the United Nations, they are “weapons that locate, select, and engage human targets without human supervision”. The word ‘engage’ in this definition is a euphemism for ‘kill’. I am not talking about weapons that are operated remotely by humans, such as the US Predator drone or Ukraine’s home-made grenade droppers, because these are not autonomous. Nor am I talking about anti-missile defence systems, or about the fully autonomous drones that both Russians and Ukrainians are using for reconnaissance, which are not lethal. And I am not talking about the science-fiction robots portrayed in the ‘Terminator’ films — controlled by the spooky emergent consciousness of the Skynet software system and driven by hatred of humanity — that the media often conjure up when discussing autonomous weapons. The issue here is not rogue machines taking over the world, but weapons deployed by humans that will drastically reduce our physical security.

Current AI systems exhibit all the required capabilities — planning missions, navigating, 3D mapping, recognizing targets, flying through cities and buildings, and coordinating attacks. Lots of platforms are available. These include: quadcopters ranging from centimetres to metres in size; fixed-wing aircraft (from hobby-sized package-delivery planes and full-sized, missile-carrying drones to ‘autonomy-ready’ supersonic fighters, such as the BAE Systems Taranis); self-driving trucks and tanks; autonomous speedboats, destroyers and submarines; and even skeletal humanoid robots.

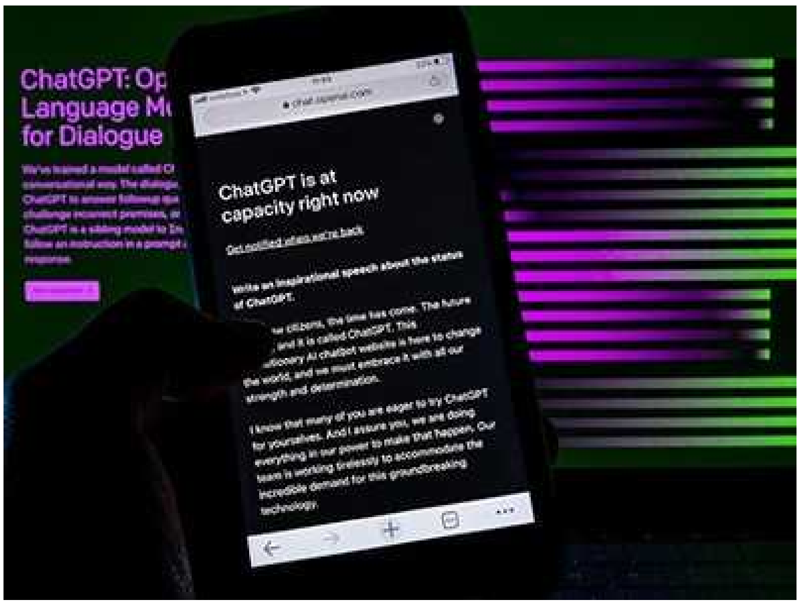


Ukrainian soldiers operate a surveillance drone on the front line near Kherson, Ukraine.Credit: Jim Huylebroek/New York Times/Redux/eyevine

The road to full autonomy in the Russia–Ukraine conflict begins with various types of semi-autonomous weapon already in use. For example, Russia is deploying ‘smart’ cruise missiles to harsh effect, hitting predefined targets such as administrative buildings and energy installations. These weapons include Iranian Shahed missiles, nicknamed ‘mopeds’ by the Ukrainians owing to their sound, which can fly low along rivers to avoid detection and circle an area while they await instructions. Key to these attacks is the use of swarms of missiles to overwhelm air-defence systems, along with minimal radio links to avoid detection. I have heard reports that new Shaheds are being equipped with infrared detectors that enable them to home in on nearby heat sources without requiring target updates communicated from controllers by radio — if true, this would be an important step towards full autonomy.

The Ukrainians have deployed Turkish Bayraktar teleoperated weapons against tanks and other targets since the early days of the war. Improved Russian air defences and jamming have made these weapons more

vulnerable and less effective over time; moreover, they cost around US$5 million each (250 times more expensive than Shaheds). Commercial, remote-controlled quadcopters that have been adapted to drop grenades have proved effective in small-scale tactical operations, and remotely piloted boats have been used to attack naval targets. But, as jamming systems become the norm, teleoperation becomes more difficult and autonomous weapons increasingly attractive.



ChatGPT: five priorities for research

Elsewhere, lethal autonomous weapons have been on sale for several years. For example, since 2017, a government-owned manufacturer in Turkey (STM) has been selling the Kargu drone, which is the size of a dinner plate and carries 1 kilogram of explosive. According to the company's website in 2019 (since toned down), the drone is capable of "autonomous and precise" hits against vehicles and persons, with "targets selected on images" and by "tracking moving targets" (see go.nature.com/3ktq6bb). As reported by the UN, Kargu drones were used in 2020 by the Libyan Government of National Accord — despite a strict arms embargo — to autonomously 'hunt down' retreating forces[1].

Other 'loitering' forms of missile, such as the Shahed, also exhibit a form of autonomy. The Israeli Harpy drone can fly over a region for several hours looking for targets that match a visual or radar signature and then destroy

them with its 23-kilogram explosive payload. (Russia's Lancet missile, widely used in Ukraine, has similar characteristics.) Whereas the Kargu and Harpy are 'kamikaze' weapons, the Chinese Ziyan Blowfish A3 is an autonomous helicopter equipped with a machine gun and several unguided gravity bombs. All of these systems are described as having both autonomous and remotely operated modes, making it difficult to know whether any given attack was carried out by a human operator.

Benefits and problems

Why are militaries pursuing machines that can decide for themselves whether to kill humans? Like remotely operated weapons, autonomous aircraft, tanks and submarines can carry out missions that would be suicidal for people. They are cheaper, faster, more manoeuvrable and have longer range than their crewed counterparts; can withstand higher *g*-forces in flight; and function underwater without life-support systems. But, unlike remotely operated weapons, autonomous weapons can function even when electronic communication is impossible because of jamming — and can react even faster than any weapon remotely controlled by a human. AI expert Kai-Fu Lee, among others, has described autonomous weapons as the 'third revolution in warfare' after gunpowder and nuclear weapons[2].

A common argument in favour is that waging war through autonomous weapons will protect military lives, just as remotely operated weapons and cruise missiles are said to do. But this is a fallacy. The other side would have such weapons, too — and as we have seen in Ukraine, the death toll among soldiers as well as civilians is staggering.

Another point often advanced is that, compared with other modes of warfare, the ability of lethal autonomous weapons to distinguish civilians from combatants might reduce collateral damage. The United States, along with Russia, has been citing this supposed benefit with the effect of blocking multilateral negotiations at the Convention on Certain Conventional Weapons (CCW) in Geneva, Switzerland — talks that have occurred sporadically since 2014.



A fleet of Kargu drones at Turkish manufacturer STM in Ankara.Credit: Mehmet Kaman/Anadolu Agency via Getty

The case relies on two claims. First, that AI systems are less likely to make mistakes than are humans — a dubious proposition now, although it could eventually become true. And second, that autonomous weapons will be used in essentially the same scenarios as human-controlled weapons such as rifles, tanks and Predator drones. This seems unequivocally false. If autonomous weapons are used more often, by different parties with varying goals and in less clear-cut settings, such as insurrections, repression, civil wars and terrorism, then any putative advantage in distinguishing civilians from soldiers is irrelevant. For this reason, I think the emphasis on the weapons' claimed superiority in distinguishing civilians from combatants, which originates from a 2013 UN report[3] pointing to the risks of misidentification, has been misguided.

There are many more reasons why developing lethal autonomous weapons is a bad idea. The biggest, as I wrote in *Nature* in 2015[4], is that "one can expect platforms deployed in the millions, the agility and lethality of which

will leave humans utterly defenceless". The reasoning is illustrated in a 2017 YouTube video advocating arms control, which I released with the Future of Life Institute (see go.nature.com/4ju4zj2). It shows 'Slaughterbots' — swarms of cheap microdrones using AI and facial recognition to assassinate political opponents. Because no human supervision is required, one person can launch an almost unlimited number of weapons to wipe out entire populations. Weapons experts concur that anti-personnel swarms should be classified as weapons of mass destruction (see go.nature.com/3yqjx9h). The AI community is almost unanimous in opposing autonomous weapons for this reason.

Moreover, AI systems might be hacked, or accidents could escalate conflict or lower the threshold for wars. And human life would be devalued if robots take life-or-death decisions, raising moral and justice concerns. In March 2019, UN secretary-general António Guterres summed up this case to autonomous-weapons negotiators in Geneva: "Machines with the power and discretion to take lives without human involvement are politically unacceptable, morally repugnant and should be prohibited by international law" (see go.nature.com/3yn6pqt). Yet there are still no rules, beyond international humanitarian laws, against manufacturing and selling lethal autonomous weapons of mass destruction.

Political action at a standstill

Unfortunately, politics has not kept up with technological advances. Dozens of human-rights and arms-control organizations have joined the Campaign to Stop Killer Robots, which calls for a ban on lethal autonomous weapons. Politicians and governments have failed to act, despite polls suggesting broad public support for such a ban (more than 60% of adults; see, for example, go.nature.com/416myef). Thousands of researchers and leaders in AI, including me, have joined those calls (see go.nature.com/4gqmfm5), yet, so far, no academic society has developed a policy on autonomous weapons because of concerns about discussing matters that are not purely scientific.



[Stop the emerging AI cold war](go.nature.com)

One reason that negotiations under the CCW have made little progress is confusion, real or feigned, about technical issues. Countries still argue endlessly about the meaning of the word 'autonomous'. Absurdly, for example, Germany declared that a weapon is autonomous only if it has "the ability to learn and develop self-awareness". China, which ostensibly supports a ban on autonomous weapons, says that as soon as weapons become capable of autonomously distinguishing between civilians and soldiers, they no longer count as autonomous and so wouldn't be banned. The United Kingdom has pledged never to develop or use lethal autonomous weapons, but keeps redefining them so that its pledge is effectively meaningless. For example, in 2011, the UK Ministry of Defence wrote that "a degree of autonomous operation is probably achievable now", but in 2017 stated that "an autonomous system is capable of understanding higher-level intent". Michael Fallon, then secretary of state for defence, wrote in 2016 that "fully autonomous systems do not yet exist and are not likely to do so for many years, if at all", and concluded that "it is too soon to ban something we simply cannot define" (see go.nature.com/3xrztn6).

Further progress in Geneva soon is unlikely. The United States and Russia refuse to allow negotiations on a legally binding agreement. The United States worries that a treaty would be unverifiable, leading other parties to circumvent a ban and creating a risk of strategic surprise. Russia now

objects that it is being discriminated against, because of its invasion of Ukraine.

A pragmatic way forward

Rather than blocking negotiations, it would be better for the United States and others to focus on devising practical measures to build confidence in adherence. These could include inspection agreements, design constraints that deter conversion to full autonomy, rules requiring industrial suppliers to check the bona fides of customers , and so on. It would make sense to discuss the remit of an AI version of the Organization for the Prohibition of Chemical Weapons, which has devised similar technical measures to implement the Chemical Weapons Convention. These measures have neither overburdened the chemical industry nor curtailed chemistry research. Similarly, the New START treaty between the United States and Russia allows 18 on-site inspections of nuclear-weapons facilities each year. And the Comprehensive Nuclear-Test-Ban Treaty might never have come into existence, had not scientists from all sides worked together to develop the International Monitoring System that detects clandestine violations.



The aftermath of a ‘kamikaze’ drone attack in Kyiv.Credit: Ed Ram/Guardian/eyevine

Despite the impasse in Geneva, there are glimmers of hope. Of those countries that have stated a position, the vast majority favours a ban. Negotiations could progress in the UN General Assembly in New York City, where no country has a veto, and at ministerial-level meetings. Last week, the government of the Netherlands hosted a meeting in The Hague on ‘responsible AI in the military domain’, where the question of whether it is ethical to introduce this class of weapon at all was raised. During the meeting, the United States announced a “political declaration” of principles and best practices for the military use of AI and urged other nations to sign up to these (see go.nature.com/3xsj779). Perhaps the most important is the statement that: “States should maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment.” Already, more than 60 countries, including China, have joined the declaration. Unfortunately, it is non-binding and does not rule out any category of autonomous weapon.

On 23–24 February, Costa Rica is due to host a meeting of Latin American and Caribbean nations on the ‘social and humanitarian impact of autonomous weapons’, which includes threats from non-state actors who might use them indiscriminately. These same nations organized the first nuclear-weapon-free zone, raising hopes that they might also initiate a treaty declaring an autonomous-weapon-free zone.

Next steps

In my view — and I suspect that of most people on Earth — the best solution is simply to ban lethal autonomous weapons, perhaps through a process initiated by the UN General Assembly. Another possibility, suggested as a compromise measure by a group of experts (see go.nature.com/3jugzxy) and formally proposed to the international community by the International Committee of the Red Cross (see go.nature.com/3k3tpan), would ban anti-personnel autonomous weapons. Like the St Petersburg Declaration of 1868, which prohibited exploding ordnance lighter than 400 grams, such a treaty could place lower limits on

the size and payload of weapons, making it impossible to deploy vast swarms of small devices that function as weapons of mass destruction.



Ethics of artificial intelligence

Instead of blocking progress in Geneva, countries should engage with the scientific community to develop the technical and legal measures that could make a ban on autonomous weapons verifiable and enforceable. Technical questions include the following. What physical parameters should be used to define the lower limit for allowable weapons? What are ‘precursor’ platforms (which can be scaled up to full autonomy), and how should their production and sale be managed? Should design constraints be used, such as requiring a ‘recall’ signal? Can firing circuits be separated physically from on-board computation, to prevent human-piloted weapons from being converted easily into autonomous weapons? Can verifiable protocols be designed to prevent accidental escalation of hostilities between autonomous systems?

On the civilian side, professional societies in AI and robotics (including the Association for the Advancement of Artificial Intelligence, the Association for Computing Machinery and the Institute of Electrical and Electronics Engineers) should develop and enforce codes of conduct proscribing work on lethal autonomous weapons. There are many precedents: for example, the American Chemical Society has a strict chemical-weapons policy (see

go.nature.com/3yn8ajt) and the American Physical Society asks the United States to ratify the Comprehensive Nuclear-Test-Ban Treaty (see go.nature.com/3jrajvr), opposes the use of nuclear weapons against non-nuclear states (see go.nature.com/3k4akq8) and advocates robust research programmes in verification science and technology for the benefit of peace and security (see go.nature.com/3hzjkkv).

As Russia's war in Ukraine unfolds, and as autonomous-weapons technology races ahead (along with the desire to use it), the world cannot afford another decade of diplomatic posturing and confusion. Governments need to deliver on what seems a simple request: to give their citizens some protection against being hunted down and killed by robots.

Nature **614**, 620-623 (2023)

doi: https://doi.org/10.1038/d41586-023-00511-5

References

1. United Nations. *Final Report of the Panel of Experts on Libya Established Pursuant to Security Council Resolution 1973 (2011)* (UN Security Council, 2021).

2. Lee, K.-F. 'The third revolution in warfare' *The Atlantic* (11 September 2021).

3. UN General Assembly. *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* (United Nations, 2013).

4. Russell, S. *Nature* **521**, 415–418 (2015).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00511-5

Work

- **Couples and colleagues: how scientist duos marry work and home life** [14 Feb 2023]
  • You can work successfully with a romantic partner. Meet the researchers who have cracked it.
- **Innovative technologies crowd the short-read sequencing market** [20 Feb 2023]
  • With a dizzying range of strategies available, laboratories must weigh up their options to find the best fit for their projects
- **Tweezers, water, mulberry paper: my toolkit for repairing ancient Japanese art** [20 Feb 2023]
  • Art conservator Jiro Ueda uses traditional techniques and natural solutions to repair priceless paintings and artefacts.

- CAREER FEATURE
- 14 February 2023

Couples and colleagues: how scientist duos marry work and home life

You can work successfully with a romantic partner. Meet the researchers who have cracked it.

- Rachael Pells

Find a new job



Pollution researchers Steve and Deonie Allen often work together on one project for one salary, which they find gets the work done in the most effective way.Credit: Trace Ward/Zephyrs

Steve and Deonie 'Dee' Allen's fates were sealed when their respective dogs, two kelpies, spotted each other from across a marina in Brisbane, Australia and became friends. The pair were living on boats three berths down from one another — and the rest, as they say, is history.

Twenty-two years later they are happily married, with one boat, two PhDs and parallel careers as microplastic-pollution researchers at Dalhousie University in Halifax, Canada, and the University of Birmingham, UK, respectively. The Allens, as many in the field will attest, are an infamous double-act, sharing work and hobbies, and travelling the world together for their sailing hobby and for research. They literally finish each other's sentences — and e-mails — and they are indubitable proof that yes, it really is possible to work with a romantic partner without growing to despise them.

In the Allens' case, Steve says, "It means we are a complete mobile research unit."

The power of two

Romantic duos are by no means uncommon in science. A 2008 study by researchers at Stanford University in California suggests that around 40% of women and 34% of men in academia are in a partnership with fellow academics[1]. Among scientists, 83% of women and 54% of men in academic couples are partnered with another scientist[1].

There are clear benefits: a support system and a shared understanding of all work's stresses, including the highs and lows of a career in science, are implicit. And couples who collaborate report higher productivity levels[2]. "The fact we share the same project goals and workload allows each of us the chance to explore new methods or equipment while the work progresses, which is priceless in order to keep up with the field," says Steve Allen.

When the COVID-19 pandemic hit in 2020, Justine Ammendolia and Jackie Saturno were able to turn a period of travel restrictions and cancelled plans into a creative, collaborative project. Stuck at home in Toronto, Canada, Ammendolia and Saturno, both environmental scientists, quickly noticed the growing number of face masks littered around the community. “It became our weird idea of a date night, walking around the community picking up garbage,” jokes Saturno, who is currently a network manager at an eco-forestry programme in Halifax, Canada, called the Family Forest Network.



Environmental scientists Justine Ammendolia (left) and Jackie Saturno work together on a research project to monitor plastic waste in Halifax, Canada.Credit: Justine Ammendolia and Jackie Saturno

The couple hatched a plan to monitor plastic waste in the area, turning it into research. “It began as a grass-roots, informal project, but we eventually picked up funding dollars,” says Ammendolia, a PhD student at Dalhousie University. From there, they started a consultancy company and were

contracted by the federal government to produce a report on the scale of the issue[3].

Ammendolia and Saturno met as undergraduates at the University of Guelph in Canada, around 12 years ago, and have worked together in both official and unofficial capacities ever since. “Our strengths play off each other,” says Ammendolia. “Scientifically, we’re pretty streamlined. Emotionally, we balance each other: Jackie is more laid back, calm and collected, and I will admit to being a bit of a firecracker.”

Key to their success in working together, Saturno thinks, is a willingness to “put egos to one side”. “Otherwise, things can unravel,” she says. “We’re always checking each other’s balances, but in a respectful way. We like to make sure that we have a sound project and we’re not just each other’s ‘yes’ person.”

Two bodies, one goal

Like many couples in science, Meaghan Creed and Alexxai Kravitz met at a conference. Both are now associate professors at Washington University School of Medicine in St. Louis, Missouri. “We were in the same scientific circles and interested in similar research questions,” says Creed. “It was great to meet someone who thought and cared deeply about the same questions that I spent a significant proportion of my time thinking and caring about.”



[‘Trailing spouses’ and ‘two body’ problems: how to move labs as a scientist couple]

In 2019, the couple were recruited to the same institution, and their laboratories are now adjacent. As a result, they frequently undertake collaborative projects, and they share a lab manager and meeting schedule.

Getting to this secure point in their careers wasn’t easy, however. When applying for positions early on, Creed says, she was advised to keep quiet about the ‘two-body problem’ in interviews and wait until she had been offered a job to disclose that her partner is also an academic.

“This is great advice, but didn’t work for us, and just led to a lot of disappointment and frustration,” she says. Creed and Kravitz were seeking two tenure-track positions, and many public institutions couldn’t offer a second such position without a lengthy approval process. It would also require an institution to conduct another full candidate search.

Although Kravitz was willing to take a step back from the tenure track to maximize the likelihood of them finding two jobs together, “that ended up not being necessary”, says Creed. On reflection, she thinks “it helps to be flexible and to emphasize the value you bring together”. “I think we got lucky,” Kravitz adds.

Luck, in many cases, seems highly dependent on the attitudes of those doing the hiring. When Creed and Kravitz were looking for job opportunities in the same city, the dean of Washington University School of Medicine heard through word of mouth that the couple were ideally looking for a career ‘package deal’. “He saw the opportunity to make two great hires for two open positions in different departments,” says Creed. “There are at least five other couples in our research sphere here at Washington University, so I think the institute sees attracting scientific couples as a strength.” But not many institutions take such a proactive approach.



One scientist couple’s five suggestions to solve the ‘two body problem’

The two-body problem comes with well-documented challenges. For one, finding the right position in the right location can be twice as difficult when you share a speciality and are effectively competing with your own partner for roles. “Scientific funding does not typically support couples working together,” says Dee Allen. “Because we are so close in our subject matter, there’s normally only ever one job going, not two.”

The Allens’ solution has been to pitch themselves to prospective employers or grant funders as a two-for-one package that produces what they think equates to the work of three people, owing to the long hours they rack up together, with inevitable late-night and early-morning brainstorming sessions.

The result is that they will often end up working together on one project for one salary — something they don't necessarily recommend to others. "But it works for us," says Dee. "It's also much more fun researching together and allows us to get the work done in the most effective way."

At the Yale School of Medicine's Cardiovascular Research Center in New Haven, Connecticut, five pairs of spouses work together across adjoining labs run by John Hwa and Kathleen Martin, who is a co-director of the centre. The phenomenon started a decade ago, when Martin and Hwa — who have been married for 28 years and colleagues for 23 — made the decision to hire their postdoc's wife alongside him, in a bid to save the couple from moving elsewhere for work.

Martin and Hwa's first couple hire fared so well that they extended similar offers to other couples whenever the opportunity arose. "When couples are able to work together in the same department or even lab, it helps to ensure the success of both individuals," says Martin. "In addition to reading and providing feedback on one another's work, we can also provide support in the face of stressful deadlines, challenging critiques or other work-related issues."



Kathleen Martin and John Hwa's research groups include five pairs of spouses working together across adjoining labs at Yale University in New Haven, Connecticut.Credit: Vivian Gu

Research suggests that hiring couples brings clear advantages: offering dual positions can help to attract star talent, compensate for jobs being in less desirable locations and, importantly, help labs to attract a diverse pool of faculty members[2].

In the 2008 study[1], Stanford University researchers concluded that failure to accommodate couples at an institution can lead to avoidable recruitment challenges. In a survey as part of the study, 88% of respondents who were hired as part of a couple said they would have gone elsewhere if their partner had not also been taken on.

Challenging assumptions

The often-precarious nature of academia is one reason partners should be prepared to stay flexible. When Ammendolia received opportunities to study in Greenland in 2014 and on the remote Fogo Island off the coast of Newfoundland in Canada in 2016, her relationship with Saturno was tested: "I was really surprised when Jackie said she would wait for me," she says. Now, the couple agree it goes without saying they would support each other from a distance once again if the right job came up.

As a queer couple, Ammendolia and Saturno say they feel fortunate to live in Canada where rights for LGBT+ people (those from sexual and gender minorities) are protected and people are welcoming, on the whole. When travelling to less tolerant countries or regions, however, Saturno notes: "It's unfortunate but we have to present as straight best friends, just because it's safer and we don't know how we would be perceived otherwise."

"It's a constant dance of figuring out who you're with, what the landscape looks like and what the social norms are there," Ammendolia adds. "Just because you're in a progressive country doesn't mean everyone holds those values."

It's clear that more could be done to protect queer academics and other people from under-represented groups when conducting fieldwork, says Saturno. "When you're sent out to do fieldwork, there's no university guidebook or safety protocol that's handed to you to advise how to survive that. We do hear comments, and often worry about safety. A lot of it's up to your intuitive nature to figure out." She would like to see universities implement formal guidance for safety protocols as a first step towards helping LGBT+ researchers to navigate fieldwork in challenging or unfamiliar places.



One couple, two cities: How to handle an international career move

Saturno and Ammendolia have learnt from their own experiences of vulnerability, and say this has made them more sensitive and empathetic towards research participants when conducting social-science studies in the field. “It encourages us to think about the things we’re asking of people,” says Ammendolia. For example, “what benefits does this project have, and what betterment are we bringing to the world? Does it risk bringing harm to the communities we’re working with? These are important questions to keep asking ourselves.”

The fact that the Allens work together so closely makes it more obvious when discrimination and sexism rear their heads. “Because Steve is a white, older male, everybody assumes that he’s my boss or my professor,” says Dee, who is a younger, petite, white woman, and is often mistakenly assumed to be less experienced. At times when Steve says he can “get on [with a project] without anyone questioning me”, Dee is asked to “fill in every single form going”. Steve is frequently paid more than Dee and given the benefit of the doubt.

“We battle it wherever possible and do our best to be strong advocates for getting women and other under-represented groups heard in science,” says Dee.

Betting on boundaries

Many people agree that retaining a level of independence in any partnership is crucial: “Don’t be competitive with each other or so collaborative to the point of poorly integrating or not having bandwidth to collaborate with others,” warns Creed.

It would be easy to advise couples to separate their work and home life, by putting a ban on science chat at the dinner table, for instance. But every scientist who is passionate about their subject knows that’s a hollow goal.

“We talk about work a lot,” admits Creed, but she views this as a positive. “Our talk at home is more about the science itself, which we love. It’s the reason we got into these jobs, but it can get easily pushed to the back-burner during the workday by other responsibilities.”

Having adjacent labs does make taking lunch breaks together easy, she adds. “It sounds silly, but it’s a built-in face time where we’re not running around multitasking.”

Debriefing with a partner at home about the day’s work can be a fun stress reliever, Creed says. Fortunately, there’s an adjudicator on hand if it gets too much: “Our three-year-old daughter is becoming much more opinionated and will put her hand over our mouths and implore us to ‘Stop talking about that!’ — which helps.”

Martin says that when worries or scientific inspiration strike after hours, it’s advantageous to have a colleague at the ready. “But it’s a balancing act to make sure that there are also work-free times to focus on our family and each other,” she says. “We probably have less separation between work and home than couples who don’t work together, especially in the work-from-home era, but date nights and getting together with friends are helpful.”

Couples agree that sharing hobbies and leisure activities can help to keep the romantic spark alive. “Spending a lot of time together for work can easily become repetitive,” says Saturno. To combat that, the couple seek new challenges. “We are learning to restore furniture together, and we go rock climbing.” Others might consider taking up another adventure sport, such as

paragliding for the Allens: “It’s the only time we’re not thinking about work, because in the moment you just can’t,” says Dee Allen.

“It’s true we wake up talking science and go to sleep talking science,” says Steve. “But we wouldn’t have it any other way.”

Nature **614**, 795-797 (2023)

doi: https://doi.org/10.1038/d41586-023-00458-7

References

1. Schiebinger, L., Davies Henderson, A. & Gilmartin, S. K. *Dual-Career Academic Couples* (The Clayman Institute, 2008).
2. Fisher, C. D. *Ind. Organ. Psychol.* **8**, 13–18 (2015).
3. Ammendolia, J. *et al. Sci. Total Environ.* **848**, 157614 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00458-7

- TECHNOLOGY FEATURE
- 20 February 2023

Innovative technologies crowd the short-read sequencing market

With a dizzying range of strategies available, laboratories must weigh up their options to find the best fit for their projects

- Michael Eisenstein



A scientist at the US Centers for Disease Control and Prevention loads an instrument for whole-genome sequencing. Credit: CDC/Science Photo Library

There's an old saying in the field of technology: "Nobody ever got fired for buying IBM" — a reference to the company's once-ubiquitous computers. Replace IBM with Illumina, a biotechnology company in San Diego, California, and the same could be said of DNA sequencing today.

Keith Robison, a computational biologist at Ginkgo Bioworks in Boston, Massachusetts, who writes about sequencing technologies on a blog called Omics! Omics!, says that for most laboratories, Illumina "is the really safe bet out there". However, IBM's days of computer market dominance are well in the past, and Illumina now faces multiple competitors who are looking to challenge — and perhaps unseat — the current giant of the sequencing marketplace.

Researchers, naturally, are paying attention. Pedro Oliveira heads the DNA-sequencing lab at the French National Sequencing Center, also known as Genoscope, in Évry. The lab recently partnered with several big European research projects, including the European Reference Genome Atlas, which will bring in an expected workload of four genomes per week. One of Genoscope's priorities will be to increase its arsenal of Illumina instruments — but that won't be the limit of its shopping list, and Oliveira has a broad range of platforms to consider.

Some instruments use complementary approaches that generate long-sequence reads spanning thousands of nucleotides, in contrast to Illumina's 'short reads', which are typically in the 100- to 200-base range. But the past year has also seen the launch of nearly half a dozen competing short-read systems, each touting their own advantages in terms of quality, efficiency and above all, cost. "These are exciting moments that we're living in," says Oliveira, "because this is the beginning of cheap sequencing." But the range of choices can be intimidating and confusing, given that most scientists are still waiting to see the actual data, and to assess how well these platforms match their projects.

A safe bet

Illumina entered the sequencing market with the acquisition of a company called Solexa in 2007. Solexa's 'sequencing by synthesis' (SBS) technology

exploits the same machinery that manufactures DNA in living cells. A template DNA strand is read by a DNA polymerase enzyme, which sequentially tacks on nucleotides that complement the template strand.

Each of the four DNA building blocks — A, T, G and C — is coupled to a specific fluorescent colour and a 'terminator' chemical group that pauses further DNA synthesis. Sensitive optics identify the added nucleotide from the resulting fluorescence, after which the tag and terminator are removed and the cycle repeats. The whole process occurs in a wafer-like 'flow cell', in which vast numbers of DNA targets are imaged simultaneously, generating millions or even billions of short reads per run.



Method of the year: long-read sequencing

That approach has been staggeringly successful. By one estimate, more than 90% of the world's sequencing data as of 2022 were generated on Illumina machines (see go.nature.com/3abj7ng). Dozens of would-be competitors have emerged to challenge Illumina over the years, but most have fallen by the wayside — many of them memorialized in the 'NGS Necropolis' (see go.nature.com/3xwvmkt). Catharine Aquino, who oversees short-read sequencing at the Functional Genomics Center Zurich in Switzerland, attributes that success to technical expertise. "It's just that the other companies were not very reliable in terms of library prep or sequencing," she says.

Illumina’s portfolio includes both compact benchtop systems for rapid analysis of small numbers of samples, such as the iSeq Sequencing System that costs US$20,000, and the larger high-end NovaSeq 6000, which costs nearly $1 million but can churn out up to 6 trillion bases (6 terabases) of sequence — roughly 2,000 times the length of one of the human genome — every 2 days. Illumina’s new NovaSeq X family of production-scale sequencers use a redesigned flow cell that can accommodate a much greater density of sequencing reactions along with a retooled SBS chemistry and upgraded optics, according to Illumina’s chief technology officer, Alex Aravanis. The company reports that its new systems, which began shipping this year, can generate up to three times as much data per run as the previous generation NovaSeq 6000, lowering the cost of sequencing to just $200 per human genome.

An array of alternatives

Beyond human genome assembly and mutational analysis, new applications have fuelled demand for more and better short-read data at a lower cost. These include everything from epigenetics to chromosomal conformation to proteomics. Aquino estimates that 60% of her facility’s work now involves single-cell RNA-seq, a sequencing-hungry technique that profiles the gene expression of thousands or millions of individual cells. To fill that surge in demand, both start-ups and established companies have entered the ring.

One established player, MGI Tech, a Shenzhen-based spin-off company from the Chinese genomics titan BGI, offers distinctive twists on an Illumina-like SBS approach. Both MGI and Illumina use a biochemical process to generate multiple copies of every strand of template DNA on the flow-cell surface, thus boosting the fluorescent signal, but MGI’s DNBSEQ platforms use a lower-cost — albeit more labour-intensive — method that converts templates into arrays of ‘DNA nanoballs’. “The data quality is really good, and it can be much more cost-effective than Illumina,” says Ioannis Ragoussis, head of genome sciences at the McGill Genome Centre in Montreal, Canada, who has used DNBSEQ instruments in his own facility.

Of the newcomers, the G4 benchtop system from Singular Genomics in San Diego is probably most like Illumina's. But the G4 also features a flow-cell design that can make it easier to run multiple sequencing experiments simultaneously. "It's really targeted towards these smaller, more flexible projects," says Stephanie Pond, vice-president of emerging technologies at the Translational Genomics Research Institute in Phoenix, Arizona, which beta-tested the G4.

Ultima Genomics' flow cell is even more distinctive. Rather than using a sealed cartridge containing complex channels to coordinate the flow of reagents, Ultima — based in Newark, California — applies sequencing reagents to the exposed surface of a spinning disc. The resulting centrifugal force distributes these materials evenly across the disc's surface, reducing both the complexity of the flow-cell design and the amount of reagents required, and thus lowering the cost of each run. Ultima also cuts costs by using a mixture of labelled and unlabelled nucleotides rather than just the costlier labelled molecules[1]. In one study, early-access users at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, documented generally comparable performance to Illumina in single-cell gene-expression experiments[2].



Ultima Genomics’ instruments run sequencing reactions on the surface of a spinning disc.Credit: Carolyn Fong/New York Times/Redux/eyevine

Finally, there are the chemistries developed by Element Biosciences, based in San Diego, and by Pacific Biosciences (PacBio)in Menlo Park, California, for new short-read instruments. Both rely on two-stage alternatives to the standard SBS approaches, in which fluorescently labelled nucleotides are not permanently incorporated into the newly synthesized DNA but rather bind transiently to the growing strand. Once they are imaged, they are then washed away and replaced by unlabelled nucleotides.

This results in a more natural DNA synthesis process while also allowing for careful optimization of the labelling step, and both Element and PacBio — a company already well-known for its sophisticated long-read systems — highlight the accuracy of their approaches. “We’ve been seeing extremely high-quality data,” says genomics researcher Christopher Mason at Weill Cornell Medicine in New York City, who has used Element’s AVITI system to profile the effects of space flight on human physiology.

Weighing up pros and cons

Sequencers fall broadly into two categories: production-scale instruments including Illumina's NovaSeq, and smaller benchtop instruments such as Illumina's NextSeq. For now, only Illumina and MGI operate across the full spectrum; other short-read companies target specific levels of throughput.

Production-scale instruments are massive and expensive, but such throughput is essential for many large-scale genomics or single-cell RNA-seq studies, and such instruments tend to form the backbone of core sequencing facilities. Stacey Gabriel, chief genomics officer at the Broad Institute says that almost all of the sequencing done at her centre, one of the world's leading genomics facilities, uses such instruments. "We have 32 NovaSeqs, and we run them very hard," she says, adding that her team will be augmenting this capacity with new NovaSeq X instruments.

Ultima also operates in this arena with its UG 100, but aims to counter the high cost of its hardware with cheaper sequencing costs. The company claims that it has the potential to deliver complete human genome sequences for $100 — half the price of the NovaSeq X. The Broad Institute was one of the UG 100's first users, and Gabriel says that although the technology is still maturing, she sees clear opportunities to incorporate it into their workflow for whole-genome analysis and high-throughput assays such as single-cell transcriptomics.

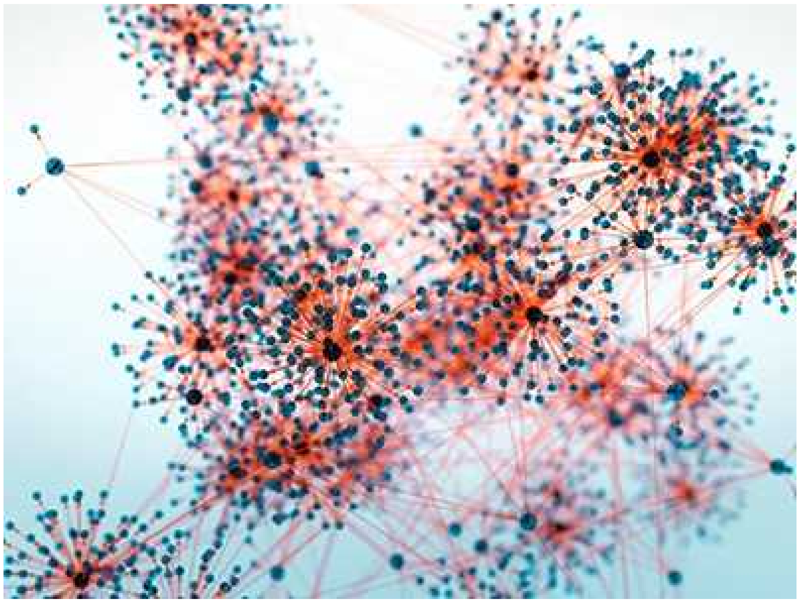


“We have 32 NovaSeqs, and we run them very hard,” says Stacey Gabriel, chief genomics officer at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts.Credit: Casey Atkins Photography

When it comes to purchasing decisions, equipment and reagents are only part of the calculation, and publicly announced per-genome prices don’t account for labour, maintenance and other support costs. Facilities can expect to pay 10% of an instrument’s base cost every year for service contracts, Ragoussis says, which can put even mid-range benchtop instruments out of reach for many labs. Most importantly, production-scale instruments are only more cost-effective relative to benchtop instruments when they are run at full capacity. “There are a lot of projects that just aren’t big enough, or pilot-scale projects where it’s really hard to ‘feed the beast’,” says Pond. This can also be an issue for labs dealing with multiple experiments that cannot be run simultaneously in a single flow cell.

Benchtop machines might be a better fit here, and this is the realm in which PacBio, Singular and Element currently compete. Such instruments generally cost between $200,000 and $400,000, and there is robust competition to deliver the most data at the lowest price-per-gigabase. “Cost

is still one of the biggest drivers, because at the end of the day people only get so much money from grants," says Mason. MGI has been using this pressure point to drive adoption of its products, Mason adds, even by offering instruments for free to some labs that are willing to spend a set amount on regular orders of reagents.



Smart software untangles gene regulation in cells

Quality is another crucial consideration, and here, too, Illumina has set a high bar. For most reads, Illumina's systems will 'call' the correct base 999 times out of 1,000 — a standard of accuracy called Q30 — and its newest-generation 'XLEAP-SBS' chemistry reportedly improves this accuracy by three-fold. PacBio claims that its new Onso instrument — which is still in beta testing — has error rates of one in 10,000 bases or lower (Q40), and Mason says his test runs with validated genomic samples have borne this out. "At the beginning of the read it's even better," he says, reporting quality nearly an order of magnitude better than Q40. Mason thinks further optimization of the computational toolbox for analysing Onso-generated data could lead to even better performance.

A 2022 preprint[3] from scientists at Element Biosciences also highlights the ability to achieve Q40 quality for most bases in a human genome sequenced with the AVITI instrument, which started shipping in June last year. The company also has a price edge over PacBio, matching Illumina's $200 cost-

per-human-genome and undercutting that of Onso by roughly sevenfold. In principle, higher quality reads reduce the amount of sequencing required for routine genomic studies and could provide a decisive advantage for applications such as the analysis of circulating tumour-derived DNA in ‘liquid biopsy’ assays. “There’s relatively few copies in the sea of normal DNA,” explains Gabriel, “so you’ve got to sequence very deeply.”

Another consideration when choosing a sequencing platform is compatibility with existing workflows. For example, Element’s workflow is largely consistent with standard Illumina processes, whereas Ultima and MGI require extra processing steps that can introduce speed bumps into existing pipelines. “It’s not insurmountable — it just adds more time and labour,” says Mason. Further automation might also be required to streamline the process.

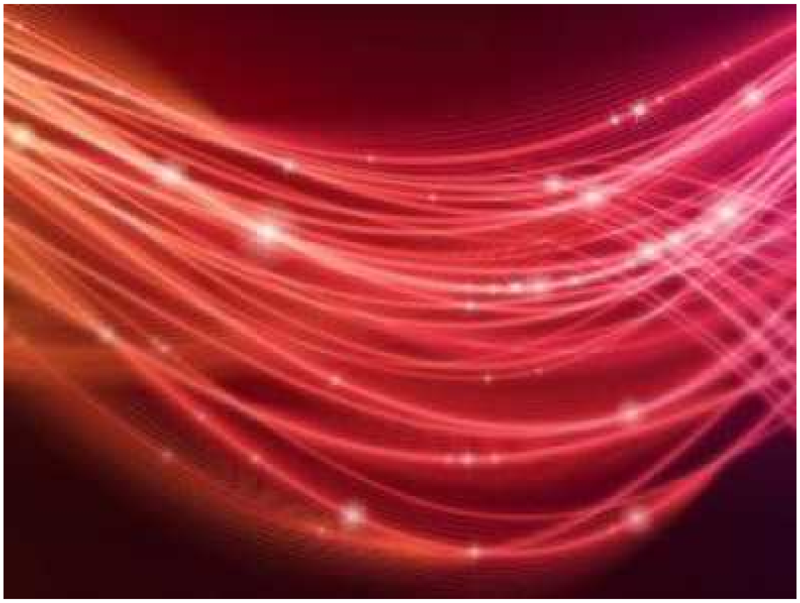
Stability and reliability are also essential, because even brief downtime can disrupt lab operations. Illumina generally has an excellent reputation on this front, says Aquino. “Sometimes even before we know something is wrong, our engineer is there already,” she says. “It will take all these companies a few more years to build up the support system and this array of experience.”

Going long

Not every sequencing application maps well onto short-read technologies. Therefore, companies such as PacBio and Oxford Nanopore Technologies (ONT) in the United Kingdom have worked to evolve their long-read technologies as well.

Both companies offer systems that directly analyse individual DNA molecules spanning tens or even hundreds of thousands of nucleotides. For PacBio, this entails feeding strands of template DNA into polymerase enzymes that are tethered to a solid surface and then using sophisticated optics to detect the addition of individual labelled nucleotides as the DNA synthesis proceeds. ONT systems determine nucleotide sequences on the basis of the distinctive changes in electrical current that occur as DNA strands transit through tiny protein pores. Collectively, these systems provide insights that would be difficult or impossible to obtain with short-

read systems, including large structural variations in chromosomal DNA, mRNA transcript structure and complete microbial genomes. Both systems can also directly identify and map epigenetic modifications.



NatureTech hub

PacBio offers some of the highest accuracy instruments on the market, thanks to a process called 'HiFi' in which the devices read the same segment of DNA over and over again, ironing out random errors along the way. However, they have historically been held back by high costs and low throughput. "A hundred samples in PacBio took a year, while 100 samples in Illumina took maybe two days," says Aquino. But the company's new Revio instrument, which costs $779,000 and is scheduled to begin shipping this year, changes the equation. With the capacity to achieve 15-fold greater throughput than current-generation systems, PacBio reports that the Revio can produce a high-quality human genome for just $1,000.

ONT offers a uniquely versatile and portable platform that can be just as easily applied to short-read applications as it can to ultra-long reads. Researchers regularly use ONT systems in the field, and Mason has even sent them to the International Space Station. "We can see applications in many remote areas," he says. ONT also offers the lowest-cost sequencing hardware on the market, including the $1,000 MinION, which can be run off of a standard laptop or, in newer versions, a tablet.

By contrast, ONT’s high-performance PromethION can sequence up to 14 terabases in 3 days, and uses an unusual business model in which most of the upfront costs are associated with the purchase of consumables needed to run sequencing experiments. “You get an instrument where the price is associated with how many flow cells you want to use without you having to buy it,” says Ragoussis, noting that this might be more appealing than spending $300,000 or more before the lab even unpacks its first flow cell. In October last year, ONT launched its newest iteration of this platform, the portable P2 Solo system, which can generate up to two human genomes per flow-cell run and allows users to get started for just over $10,000.

In such a crowded marketplace, where change is a constant, investing in new technology requires a leap of faith. “It’s very difficult to adapt every six months to a new technology — it demands a lot from the community for benchmarking, for testing and also for the bioinformatics team,” says Oliveira. For now, his team is carefully weighing up the pros and cons of these emerging platforms and how they might complement or supplant his existing hardware. But competition, in general, is a good thing, driving performance up and costs down. “We are democratizing sequencing,” he says.

Nature **614**, 798-800 (2023)

doi: https://doi.org/10.1038/d41586-023-00512-4

References

1. Almogy, G. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2022.05.29.493900 (2022).

2. Simmons, S. K. *et al. Nature Biotechnol.* **41**, 204–211 (2023).

3. Arslan, S. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2022.11.03.514117 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00512-4

- WHERE I WORK
- 20 February 2023

Tweezers, water, mulberry paper: my toolkit for repairing ancient Japanese art

Art conservator Jiro Ueda uses traditional techniques and natural solutions to repair priceless paintings and artefacts.

- Rachael Pells



Jiro Ueda is an art conservator at the Smithsonian National Museum of Asian Art in Washington DC. Credit: Stephen Voss/Smithsonian

I started working in 1996 at the Smithsonian National Museum of Asian Art in Washington DC. I came from Japan, thinking it would be interesting to stay in the United States for a couple of years. Twenty-seven years later, I am still here, and still using the same traditional techniques to repair artwork, such as this thirteenth-century Buddhist painting *The Miraculous Interventions of Jizo Bosatsu*.

In the photograph, I am using a pair of tweezers to carefully remove layers of lining paper and patches built up from centuries of repairs. Some creases have developed along the scroll that have caused the pigments to flake and powder. I can eliminate those creases by moistening the painting to make the paper expand and relax.

Sometimes ‘foxing’ — blemishes from mould spores, and other impurities — appears on the scroll. In the West, people often think these stains need to be bleached, but purified water can greatly reduce them without using chemicals.

Here, I’m taking the scroll apart; the next step is to put it back together using new patches of similar paper over the holes that have arisen with age. I will re-line the scroll with paper made from mulberry-tree fibres, using traditional wheat-starch paste as an adhesive.

Japanese art conservation is not just about the finished product, it is also about preserving the cultural heritage. The philosophy behind art conservation is the only aspect of my work that has changed in 27 years: previously, if a work was damaged, we would completely remount it. Now we do as little as possible so that we can preserve the whole object.

It is most important to find the right balance between repairing something to make it stable and functional, and protecting as much of the existing craft as possible. It takes years of experience to look at an ancient artefact such as this and understand what is necessary to repair it. I find it highly fulfilling to be part of this process.

Nature **614**, 802 (2023)

doi: https://doi.org/10.1038/d41586-023-00513-3

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00513-3

Research

- **JWST opens a window on exoplanet skies** [15 Feb 2023]
  • An unprecedented glimpse of a distant planet reveals clues about how it might have formed. Scientists explain why it's a win for atmospheric chemistry, and celebrate the technology that made it possible.
- **Identification of carbon dioxide in an exoplanet atmosphere** [02 Sept 2022]
  • Transmission spectroscopy observations from the James Webb Space Telescope show the detection of carbon dioxide in the atmosphere of the gas giant exoplanet WASP-39b.
- **Early Release Science of the exoplanet WASP-39b with JWST NIRCam** [09 Jan 2023]
  • Time-series observations from the JWST of the transiting exoplanet WASP-39b show gaseous water in the planet's atmosphere and place an upper limit on the abundance of methane.
- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM** [09 Jan 2023]
  • A broad-wavelength 0.5–5.5 µm atmospheric transmission spectrum of WASP-39b, a 1,200 K, roughly Saturn-mass, Jupiter-radius exoplanet, demonstrates JWST's sensitivity to a rich diversity of exoplanet compositions and chemical processes.
- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H** [09 Jan 2023]
  • The medium-resolution transmission spectrum of the exoplanet WASP-39b, described using observations from the Near Infrared Spectrograph G395H grating aboard JWST, shows significant absorption from CO2 and H2O and detection of SO2.
- **Early Release Science of the exoplanet WASP-39b with JWST NIRISS** [09 Jan 2023]
  • The transmission spectrum of the exoplanet WASP-39b is obtained using observations from the Single-Object Slitless Spectroscopy mode of the Near Infrared Imager and Slitless Spectrograph instrument aboard the JWST.
- **Suppressing quantum errors by scaling a surface code logical qubit** [22 Feb 2023]
  • A study demonstrating increasing error suppression with larger surface code logical qubits, implemented on a superconducting quantum processor.

- **[Less extreme and earlier outbursts of ice-dammed lakes since 1900]** [15 Feb 2023]
  • An assessment of ice-dam failures in six mountain regions shows that extreme peak flows and volumes have declined sharply since 1900, and that ice-dam floods today originate at higher elevations and earlier in the year.
- **[Severe multi-year drought coincident with Hittite collapse around 1198–1196 bc]** [08 Feb 2023]
  • Tree-ring analysis is correlated with stable isotope data and historical texts to identify an unusual multi-year period of drought from around 1198 to 1196 bc associated with the collapse of the Hittite Empire.
- **[The cellular coding of temperature in the mammalian cortex]** [08 Feb 2023]
  • A study using calcium imaging in the mouse forepaw system identifies neurons in the posterior insular cortex that respond to cooling and/or warming with distinct response dynamics.
- **[A NPAS4–NuA4 complex couples synaptic activity to DNA repair]** [15 Feb 2023]
  • A neuron-specific activity-dependent DNA repair mechanism is identified, the impairment of which may lead to neurodevelopmental disorders, neurodegeneration and ageing.
- **[Dissecting cell identity via network inference and in silico gene perturbation]** [08 Feb 2023]
  • A machine-learning-based strategy called CellOracle combines computational perturbation with modelling of gene-regulatory networks to analyse how cell identity is regulated by transcription factors, and correctly predicts phenotypic changes after transcription factor perturbation in the developing zebrafish.
- **[Influenza vaccination reveals sex dimorphic imprints of prior mild COVID-19]** [04 Jan 2023]
  • Immune responses to influenza vaccination are affected by previous mild COVID-19 in a sex-dimorphic manner.
- **[Functional T cells are capable of supernumerary cell division and longevity]** [18 Jan 2023]
  • Through iterative cycles of viral challenge and rechallenge over ten years, mouse T cells are demonstrated to have essentially infinite potential for population expansion and longevity without malignant transformation or loss of functional competence.
- **[Telomere-to-mitochondria signalling by ZBP1 mediates replicative crisis]** [08 Feb 2023]

• Dysfunctional telomeres activate innate immune responses through mitochondrial TERRA–ZBP1 complexes to eliminate cells that are destined for neoplastic transformation.

- **De novo design of luciferases using deep learning** [22 Feb 2023]
  • A deep-learning-based strategy is used to design artificial luciferases that catalyse the oxidative chemiluminescence of diphenylterazine with high substrate specificity and catalytic efficiency.
- **Structural basis for substrate selection by the SARS-CoV-2 replicase** [01 Feb 2023]
  • Cryo-EM is used to visualize the SARS-CoV-2 RTC bound to each of the natural NTPs as well as remdesivir triphosphate (RDV-TP) in states poised for incorporation, explaining the interactions required for NTP recognition and RDV-TP selectivity.

- NEWS & VIEWS FORUM
- 15 February 2023

JWST opens a window on exoplanet skies

An unprecedented glimpse of a distant planet reveals clues about how it might have formed. Scientists explain why it’s a win for atmospheric chemistry, and celebrate the technology that made it possible.

- Julia V. Seidel ,
- Louise D. Nielsen &
- Subhajit Sarkar

THE PAPERS IN BRIEF

• NASA’s James Webb Space Telescope (JWST) was launched in December 2021 and now orbits the Sun, some 1.5 million kilometres from Earth.

• Early data released last year confirm that this is an ideal vantage point for investigating exoplanets — the distant worlds that orbit stars other than the Sun.

• Five papers in *Nature* report analyses profiling the atmospheric chemistry of WASP-39b, a hot exoplanet with a Saturn-like mass.

• The studies settle questions about this exoplanet’s atmosphere, and showcase the power and versatility of JWST.

JULIA V. SEIDEL & LOUISE D. NIELSEN: Composition and origins of far-flung worlds

Rustamkulov *et al.*[1], Alderson *et al.*[2], Feinstein *et al.*[3] and Ahrer *et al.*[4,5] used three different instruments on board JWST — each with its own advantages and shortcomings — but reported largely complementary results (Fig. 1). In all five investigations, the teams found that elements heavier than hydrogen and helium are more abundant in the atmosphere of WASP-39b than they are in the Sun, whereas the ratio of carbon to oxygen is lower than that of the Sun and commensurate with that of Saturn. These findings offer crucial information about the planet's formation, the basic composition of its atmosphere and its potential to host life.

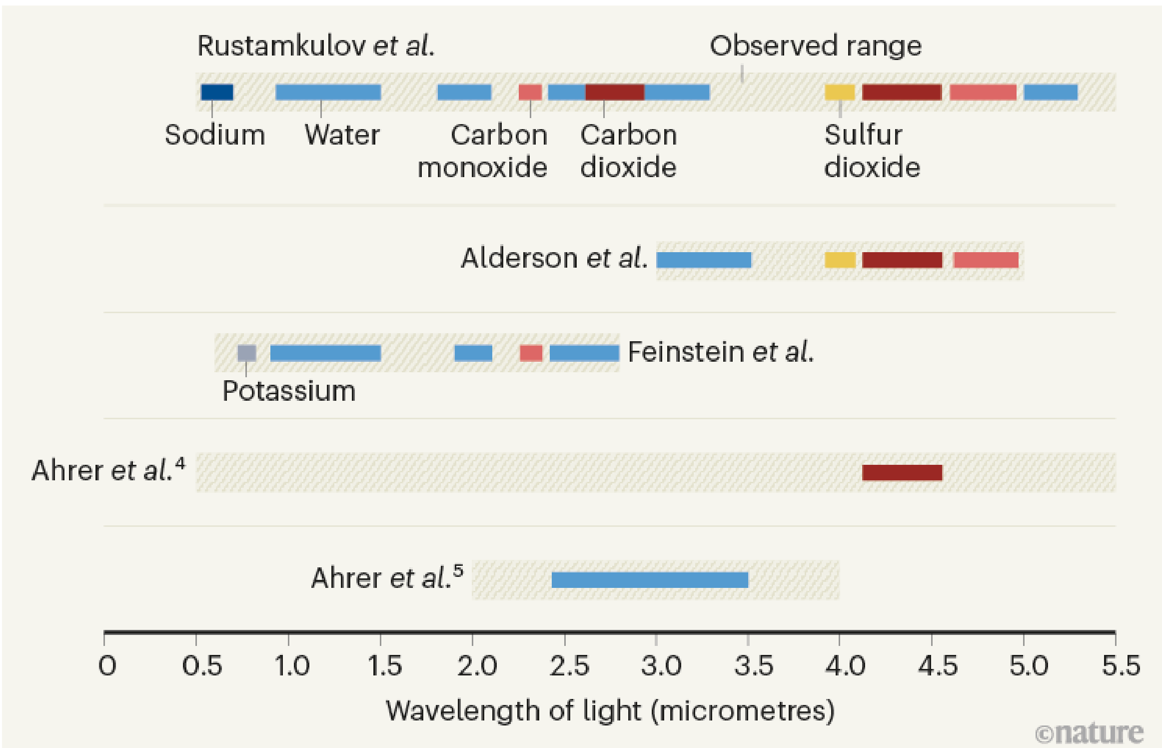
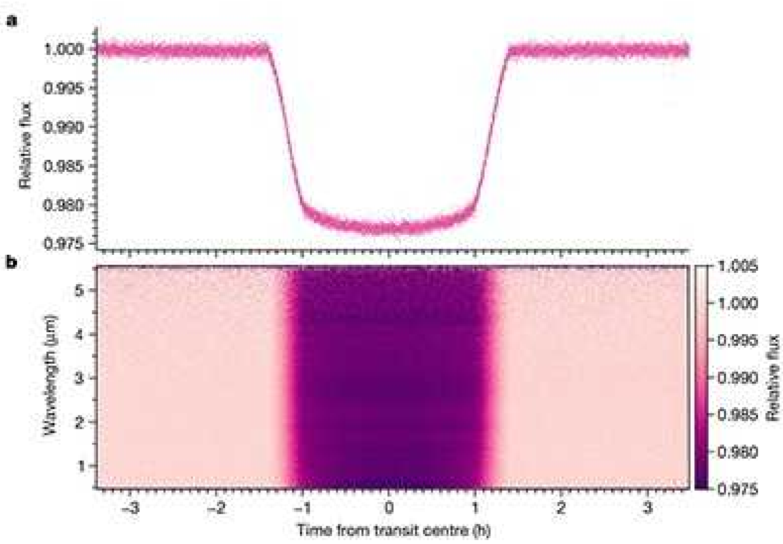


Figure 1 | Atmospheric chemistry from afar. Five papers report analyses that use the James Webb Space Telescope (JWST) to deduce the chemical composition of the atmosphere of the distant planet WASP-39b. The analyses mapped the atmosphere's transmission spectrum, which indicates how well light at various wavelengths permeates this atmosphere. The authors of each paper accessed a specific wavelength range by using one of three instruments (or configurations): Rustamkulov *et al.*[1], Alderson *et al.*[2] and Ahrer *et al.*[4] used the Near Infrared Spectrograph; Feinstein *et al.*[3] used the Near Infrared Imager and Slitless Spectrograph; and Ahrer *at al.*[5] used the Near Infrared Camera. The studies revealed signatures of water, carbon

dioxide, carbon monoxide, sodium and potassium; determined the properties of clouds; and uncovered evidence of sulfur dioxide. These findings are crucial for understanding the atmospheric processes occurring in WASP-39b and how and where the planet formed. (Adapted from graphic at go.nature.com/3ddesjg.)

The carbon/oxygen ratio of an exoplanet's atmosphere is a telltale sign of where the planet formed[6]. This is particularly useful in the case of giant planets that are close to their host stars, because their formation mechanism has been an open question since the first exoplanet was found. The ratio measured for WASP-39b indicates that the planet might have formed at a location beyond the system's water-ice line — the distance from the host star at which it is cold enough for compounds such as water and carbon dioxide to condense into solid ice. At this location, the planet could have accreted the oxygen-rich solids measured by JWST, before migrating inwards to its current position.

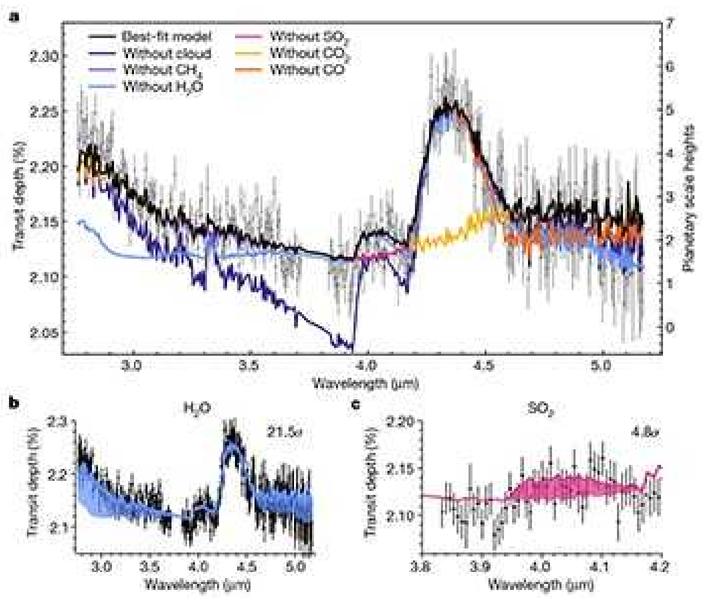


Read the paper: Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM

The sulfur/oxygen ratio is another piece in the puzzle of planetary formation. But the sulfur content of an exoplanet's atmosphere is intriguing for another reason. Sulfur dioxide is much like the protective ozone in Earth's atmosphere: it is produced during chemical reactions that are triggered by

ultraviolet radiation from the host star[7]. Rustamkulov *et al.* and Alderson *et al.* both detected sulfur dioxide in the atmosphere of WASP-39b. This observation marks the first direct evidence of light-induced (photochemical) reactions in an exoplanet atmosphere —a milestone in the quest for a truly habitable planet.

Much work remains to be done to probe the limits of this habitability. However, the finding is a step towards understanding how photochemistry protects exoplanetary surfaces from high-energy irradiation. It also tightens constraints on the parameters used in models of planetary formation. Both advances pave the way to future observations of planets that are similar to Earth.



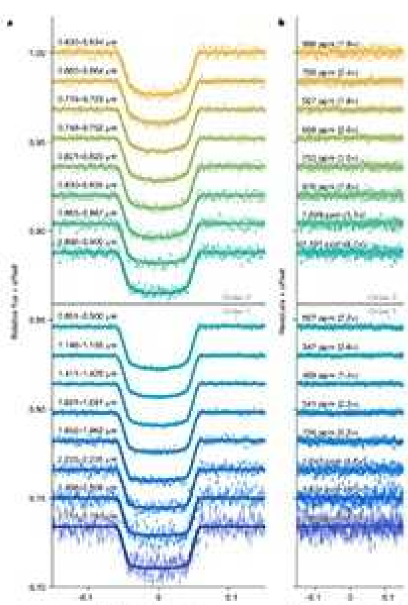
Read the paper: Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H

Part of these efforts involve profiling the characteristics of the exoplanetary atmosphere itself. By comparing the measured chemical abundances with those of several cloud models, Feinstein *et al.* determined that the clouds of WASP-39b are broken up along the day–night terminator, the line that separates day and night on a planet. Such cloud structure has previously been associated with other hot exoplanets that have masses similar to that of Jupiter[8].

Further analysis of JWST data could reveal even more information about the formation location, cloud composition and photochemistry of WASP-39b. High-resolution ground-based observations will provide crucial constraints on the chemical content of its atmosphere — its potassium and sodium, for example, and the dynamic processes with which these chemical elements are associated. Indeed, the current construction of several extremely large telescopes on Earth is promising for our understanding of alien atmospheres — because JWST's impressive sensitivity can then be married with the high spectral and spatial resolution achievable from the ground.

SUBHAJIT SARKAR: Exceptional data from an extraordinary telescope

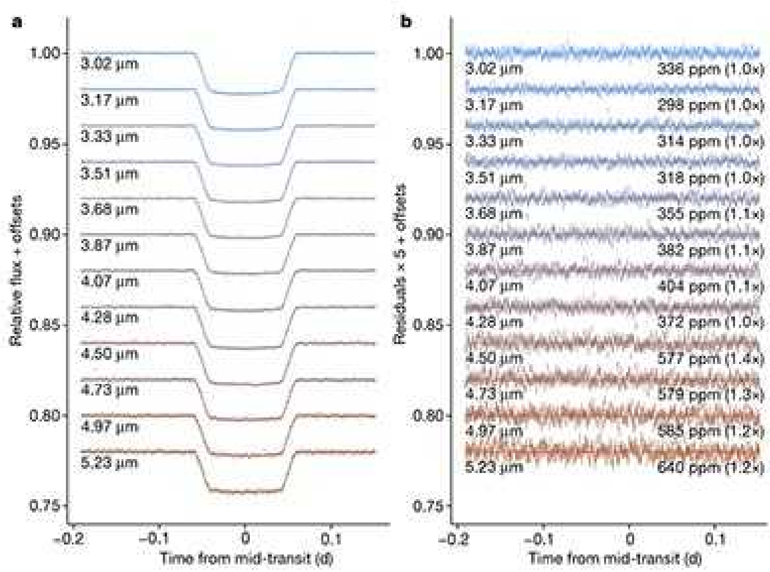
The chemistry of an atmosphere is revealed in its 'transmission spectrum', which indicates how well light at various wavelengths can permeate the gas surrounding an exoplanet. This is typically obtained using a technique known as transit spectroscopy, an approach that involves monitoring changes in the intensity of starlight at discrete wavelengths as a planet transits its host star. Although this method has previously been used to study the atmosphere of WASP-39b[9], the present papers showcase the remarkable precision and quality of the data obtained with JWST, revealing the telescope's tremendous potential — and highlight the challenges to come.



[Read the paper: Early Release Science of the exoplanet WASP-39b with JWST NIRISS]

The wavelength ranges of the three instruments used in the studies are all in the near infrared, the range in which one expects to find spectral features of the key atmospheric molecules reported. But each instrument has different configurations that enable access to different wavelength ranges and spectral resolving powers. Rustamkulov *et al.*, Alderson *et al.* and Ahrer *et al.*[4] used two configurations of an instrument called the Near Infrared Spectrograph; Feinstein *et al.* used the Near Infrared Imager and Slitless Spectrograph; and Ahrer *et al.*[5] used a device known as the Near Infrared Camera. Between them, the teams observed light with wavelengths spanning 0.5 to 5.5 micrometres (Fig. 1).

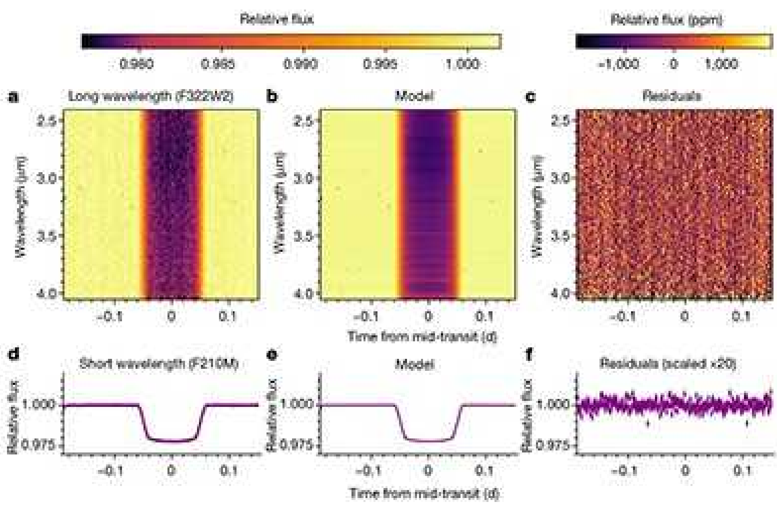
This huge wavelength coverage makes possible the extraordinary scientific results obtained by JWST and is a much larger range than that of the Hubble Space Telescope, with which previous WASP-39b spectra were obtained[9]. The spectral resolution of JWST is also more powerful than that of its predecessor. Furthermore, JWST's primary mirror has a diameter of 6.5 metres compared with Hubble's 2.4 metres (see go.nature.com/3jhjfzu) — an increase that boosts the signal-to-noise ratio, enabling the instruments' exquisite precision.



[Read the paper: Identification of carbon dioxide in an exoplanet atmosphere]

Models for exoplanetary atmospheres have long been put forward to interpret the spectra[10], but without sufficient quality of data with which to refine and constrain these models, the spectra can be interpreted in different ways. The high quality of the data reported in these studies is therefore crucial to nailing down the details of WASP-39b — including everything from how it was formed to how its clouds behave.

Various data-processing algorithms were used for each instrument to traverse the complex path from raw data to the final spectrum. And although these efforts returned mostly similar results, the diverse approaches show that there is no consensus yet on the optimal way to process JWST data — indeed, there is much still to learn. Some problems have already begun to be addressed in these initial studies, including how best to manage pixels that are saturated[1] (giving unreliable signals), and what to do when mirror segments spontaneously tilt during observation[2].



[Read the paper: Early Release Science of the exoplanet WASP-39b with JWST NIRCam]

Further solutions will be developed as our understanding of the instruments improves with time, but as a first test of JWST's performance, the results of these studies are very exciting. The findings are also encouraging for research beyond that of exoplanetary atmospheres. In particular, the WASP-39b successes show the ability of JWST to deliver data of exceptional

quality, which will enhance our understanding of the early Universe, galaxies and star formation, as well as the search for life. There is now little doubt that JWST will deliver on its promise to transform astronomy — and, more specifically, exoplanet science — in the next decade.

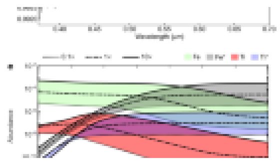
Nature **614**, 632-633 (2023)

doi: https://doi.org/10.1038/d41586-023-00394-6

References

1. Rustamkulov, Z. *et al. Nature* **614**, 659–663 (2023).
2. Alderson, L. *et al. Nature* **614**, 664–669 (2023).
3. Feinstein, A. D. *et al. Nature* **614**, 670–675 (2023).
4. JWST Transiting Exoplanet Community Early Release Science Team (Ahrer, E.-M. *et al.*) *Nature* **614**, 649–652 (2022).
5. Ahrer, E.-M. *et al. Nature* **614**, 653–663 (2023).
6. Öberg, K. I., Murray-Clay, R. & Bergin, E. A. *Astrophys. J.* **743**, L16 (2011).
7. Tsai, S.-M. *et al.* Preprint at https://arxiv.org/abs/2211.10490 (2022).
8. Parmentier, V., Fortney, J. J., Showman, A. P., Morley, C. & Marley, M. S. *Astrophys. J.* **828**, 22 (2016).
9. Wakeford, H. R. *et al. Astron. J.* **155**, 29 (2018).
10. Madhusudhan, N. *Annu. Rev. Astron. Astrophys.* **57**, 617–663 (2019).

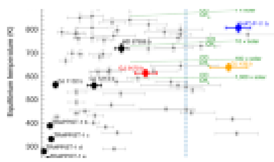
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00394-6



Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b

15 August 2018

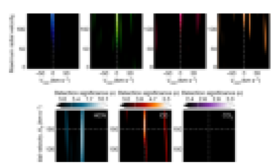
H. Jens Hoeijmakers, David Ehrenreich, … Luca Di Fabrizio



A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds

01 July 2019

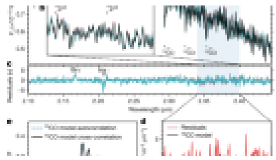
Björn Benneke, Heather A. Knutson, … Joshua Kammer



Five carbon- and nitrogen-bearing species in a hot giant planet’s atmosphere

07 April 2021

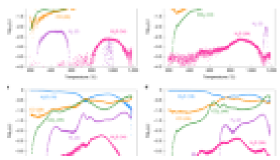
Paolo Giacobbe, Matteo Brogi, … Andrea Tozzi



[The 13CO-rich atmosphere of a young accreting super-Jupiter]

14 July 2021

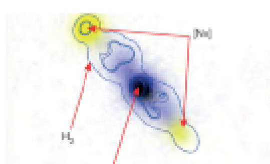
Yapeng Zhang, Ignas A. G. Snellen, … Frans Snik



[Composition of terrestrial exoplanet atmospheres from meteorite outgassing experiments]

15 April 2021

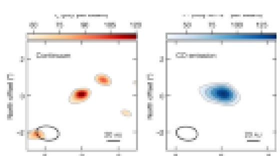
Maggie A. Thompson, Myriam Telus, … David Lederman



[Extreme 13C,15N and 17O isotopic enrichment in the young planetary nebula K4-47]

19 December 2018

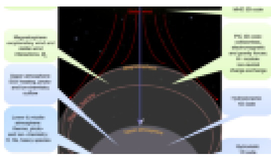
D. R. Schmidt, N. J. Woolf, … L. M. Ziurys



[Carbon monoxide gas produced by a giant impact in the inner region of a young system]

20 October 2021

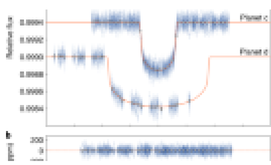
Tajana Schneiderman, Luca Matrà, … Mark C. Wyatt



[Signatures of strong magnetization and a metal-poor atmosphere for a Neptune-sized exoplanet]

16 December 2021

Lotfi Ben-Jaffel, Gilda E. Ballester, … Mercedes López-Morales



[Transit detection of the long-period volatile-rich super-Earth ν2 Lupi d with CHEOPS]

28 June 2021

Laetitia Delrez, David Ehrenreich, … Nicholas A. Walton

- Article
- Open Access
- Published: 02 September 2022

Identification of carbon dioxide in an exoplanet atmosphere

- JWST Transiting Exoplanet Community Early Release Science Team

Nature **volume 614**, pages 649–652 (2023)

- 13k Accesses
- 10 Citations
- 122 Altmetric
- Metrics details

Abstract

Carbon dioxide (CO_2) is a key chemical species that is found in a wide range of planetary atmospheres. In the context of exoplanets, CO_2 is an indicator of the metal enrichment (that is, elements heavier than helium, also called 'metallicity')[1,2,3], and thus the formation processes of the primary atmospheres of hot gas giants[4,5,6]. It is also one of the most promising species to detect in the secondary atmospheres of terrestrial exoplanets[7,8,9]. Previous photometric measurements of transiting planets with the Spitzer Space Telescope have given hints of the presence of CO_2, but have not yielded definitive detections owing to the lack of unambiguous spectroscopic identification[10,11,12]. Here we present the detection of CO_2 in the atmosphere of the gas giant exoplanet WASP-39b from transmission spectroscopy observations obtained with JWST as part of the Early Release Science programme[13,14]. The data used in this study span 3.0–5.5 micrometres in wavelength and show a prominent CO_2 absorption feature at 4.3 micrometres (26-sigma significance). The overall spectrum is well matched by one-dimensional, ten-times solar metallicity models that assume radiative–convective–thermochemical equilibrium and have moderate cloud

opacity. These models predict that the atmosphere should have water, carbon monoxide and hydrogen sulfide in addition to CO_2, but little methane. Furthermore, we also tentatively detect a small absorption feature near 4.0 micrometres that is not reproduced by these models.

Main

WASP-39b is a hot (planetary equilibrium temperature of 1,170 K assuming zero albedo and full heat redistribution), transiting exoplanet that orbits a G7-type star with a period of 4.055 days[15]. The planet has approximately the same mass as Saturn ($M = 0.28\ M_J$, where M_J is the mass of Jupiter) but is about 50% larger (radius $R = 1.28\ R_J$, where R_J is the radius of Jupiter), probably owing to the high level of irradiation that it receives from its host star[16,17,18]. We chose this planet for the JWST Early Release Science (ERS) transmission spectroscopy observations because analyses of existing space- and ground-based data detected large spectral features and showed that there was minimal contamination of the planetary signal from stellar activity[10,19,20,21]. The main spectral features previously detected were confidently attributed to sodium, potassium and water vapour absorption[10,19,20], whereas carbon dioxide (CO_2) was suggested to explain the deep transit at 4.5 μm seen with Spitzer[10].

Atmospheric metallicity has long been thought to be a diagnostic of the relative accretion of solids and gas during the formation of gas giant planets, both of which bring heavy elements to the hydrogen-dominated envelope and visible atmosphere[4,5,6]. The metallicity of WASP-39b's host star, which is a proxy for the metal enrichment of the protoplanetary disk that the planet formed in, is approximately solar[15,22,23,24]. Therefore, the planet mass–atmospheric metallicity trend observed in the Solar System giants[25,26] predicts that it has an enhancement of about ten-times solar (like that of Saturn; ref. [27]). In addition, interior structure models that match WASP-39b's low density predict a 95th percentile upper limit for the atmospheric metallicity of 55-times solar, under the limiting assumption that the planet has no heavy-element core and that all the metals are evenly distributed throughout the envelope[28].

Despite having some of the highest signal-to-noise detections of spectral features in its transmission spectrum, modelling of the existing data for WASP-39b has resulted in metallicity estimates ranging across five orders of magnitude, from 0.003-times solar to 300-times solar[10,29,30,31,32,33]. The wide range of values stems from the data being of insufficient quality to break the degeneracy between clouds and metallicity in transmission spectra models[34], as well as uncertainty over the interpretation of the photometric measurements by the Spitzer Space Telescope at 3.6 µm and 4.5 µm. Thus, spectroscopic data with greater precision, finer spectral channels and wider wavelength coverage were needed to better constrain the metallicity of this (and other) giant exoplanet atmospheres.

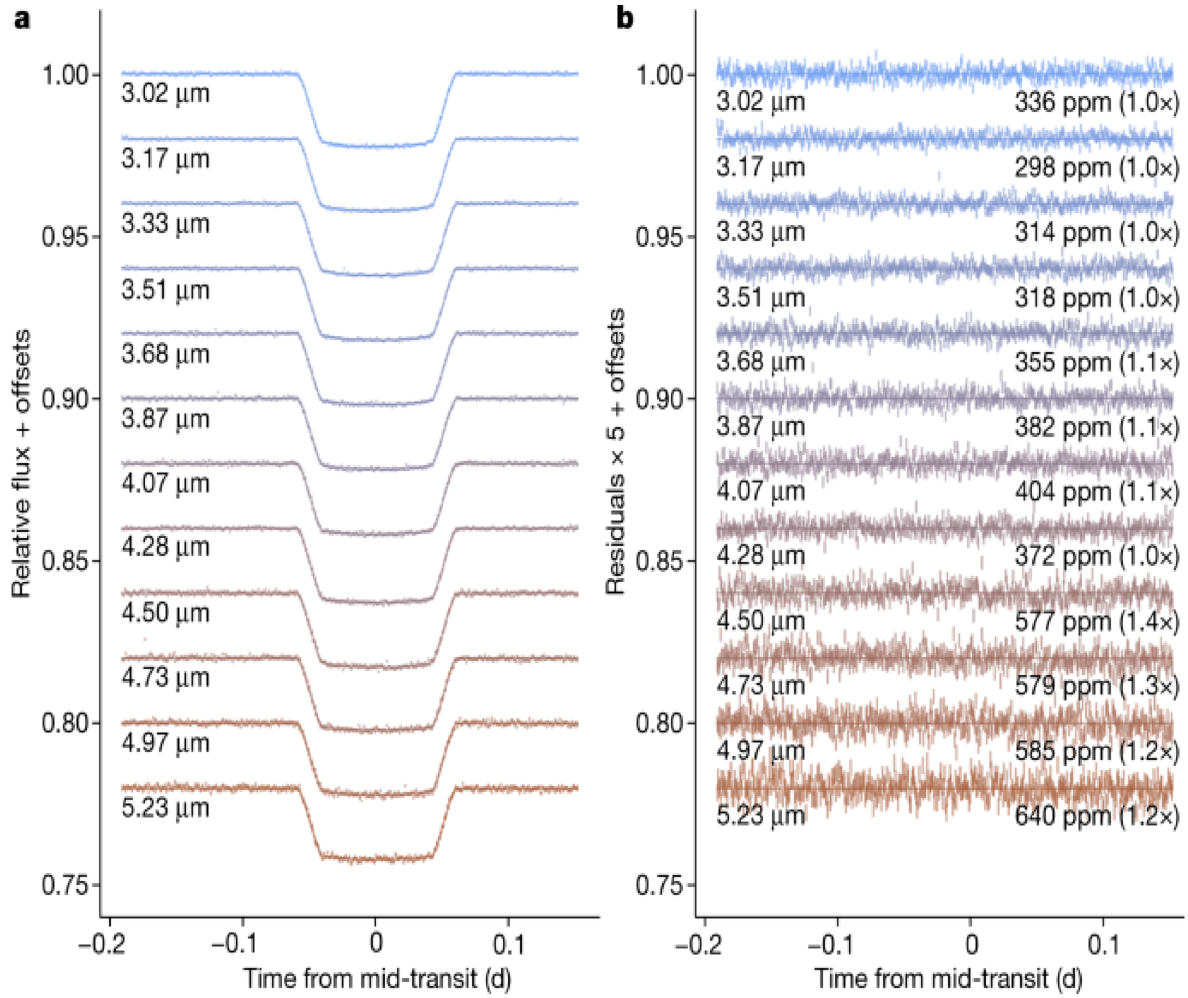
The first JWST ERS observation of WASP-39b was obtained using the Near Infrared Spectrograph (NIRSpec)[35,36] on 10 July 2022, between 15:24 and 23:37 utc. We used the Bright Object Time Series (BOTS) mode with the 1.6″ × 1.6″ fixed-slit aperture and the PRISM disperser to capture spectra between 0.5 µm and 5.5 µm. The data were recorded using the SUB512 subarray with five groups per integration and the NRSRAPID readout pattern, which gave integration times of 1.38 s. NIRSpec obtained a total of 21,500 integrations over 8.23 h of observations centred on the 2.8-h transit duration of WASP-39b.

The count rate in the PRISM mode varies significantly over the bandpass owing to the spectral energy distribution of the star and the wavelength dependency of the spectrograph dispersion. Therefore, the observations were designed to saturate at shorter wavelengths in order to obtain sufficient signal-to-noise ratio at the longer wavelengths in the bandpass that have not previously been studied spectroscopically. Wavelengths between 0.71 µm and 2.09 µm have at least one group saturated in the pixel at the centre of the spectral trace. We concentrate here on the analysis of the data longwards of 3.0 µm that are not impacted by saturation to investigate the spectrum overlapping with the previous 3.6 µm and 4.5 µm Spitzer photometric measurements. The subset of the PRISM data described herein has a native spectral resolving power ($R = \lambda/\Delta\lambda$, where λ is wavelength) of 100–350. For this study, we binned the data to lower resolving powers (values range from 60 to 200 depending on wavelength and reduction). The binning is done at the light-curve level before the fitting of the transit depths that constitute the

transmission spectrum. Analyses of JWST/NIRSpec transit observations obtained during commissioning have shown that similar levels of binning as we use here results in minimal systematics[37]. An analysis of the complete PRISM dataset at full resolution, including recovery of the saturated part of the spectrum, is ongoing.

We reduced the NIRSpec PRISM data for WASP-39b using the JWST Science Calibration Pipeline along with customized routines to minimize noise in the time-series spectra (Methods). We performed four different reductions of the transmission spectrum starting from the uncalibrated data[21,38,39,40]. Figure 1 shows derived spectroscopic transit light curves from one of the reductions. We confirm with our analysis of the WASP-39b data that NIRSpec transit observations at a resolving power of 60–200 are nearly free of systematics. We achieved close to photon-noise-limited measurements in the spectroscopic light curves after trimming the first 10 min of data and removing a linear trend in time with an average rate of about 190 ppm h^{-1} across the bandpass. We also obtained similar results by fitting the full time series with a downwards trending exponential ramp (timescale about 100 min) combined with a quadratic function of time. The lack of large systematics in these data stands in contrast to previous transit spectroscopy observations with space- or ground-based telescopes[41].

Fig. 1: JWST NIRSpec time-series data for WASP-39b.



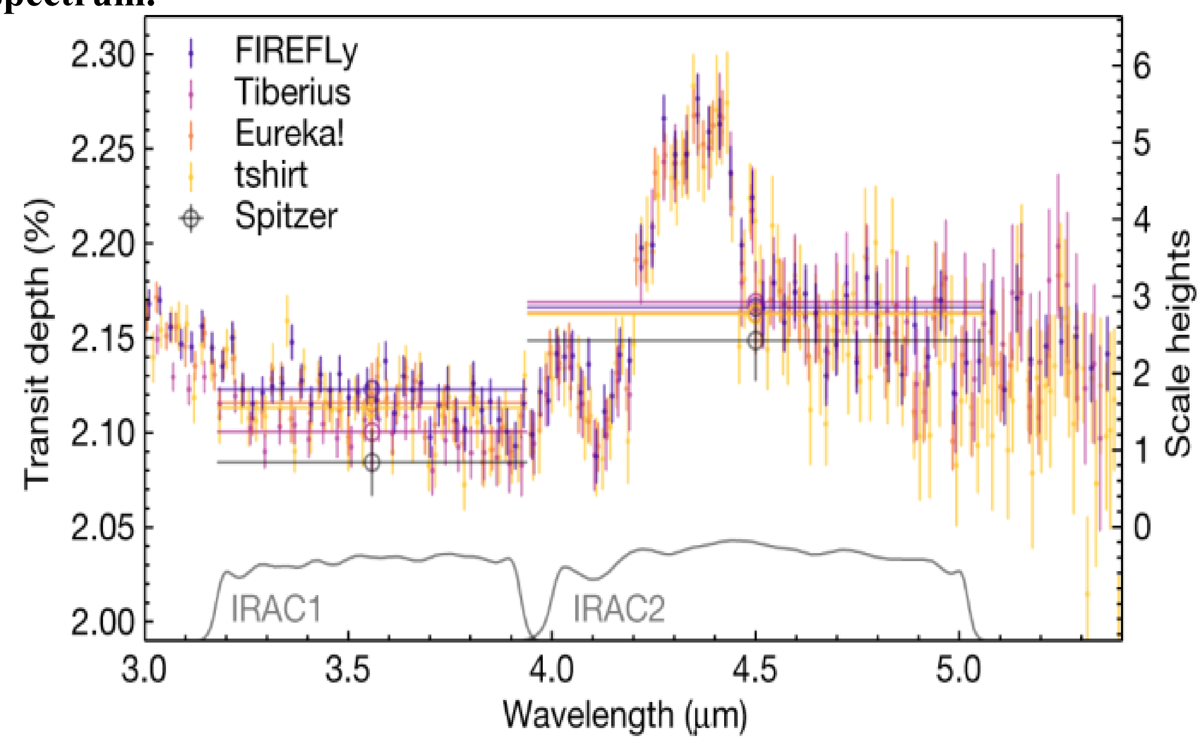
a, Spectroscopic light curves for WASP-39b's transit with a spectral resolving power of 20 and a time cadence of 1 min (data are binned and offset vertically for display purposes only). An exoplanet light-curve model was fitted to the data using a quadratic limb-darkening law with an exponential ramp and a quadratic function of time removed. **b**, Residuals of the binned light curve after subtracting the transit model scaled up by a factor of five to show the structure. The r.m.s. of the residuals are given in units of ppm. The numbers in brackets are the ratio of the r.m.s. to the predicted photon-limited noise.

Source data

The transmission spectra derived from the different reductions, shown in Fig. 2, have excellent agreement. They all show a large feature at 4.3 μm, as well as a smaller feature near 4.0 μm (discussed below). Detailed modelling

of the Fast InfraRed Exoplanet Fitting Lyghtcurve (FIREFLy)-reduced data yields a statistical significance of 26σ for the large feature (Methods). We attribute this feature to CO_2 absorption based on a comparison of the resolved band shape with theoretical models and the spectra of brown dwarfs[42]. Figure 2 also includes Spitzer's two broadband photometric measurements[10], which are consistent with the JWST data to better than 2σ after integrating the transmission spectrum over the Spitzer bandpasses. We also see good agreement (better than 2σ for all reductions) in the relative transit depths between the 3.6-µm and 4.5-µm channels. The comparison shown in Fig. 2 demonstrates both the consistency in the derived spectra from multiple, independent analyses and the reliability of the previous Spitzer measurements.

Fig. 2: Independent reductions of the WASP-39b transmission spectrum.

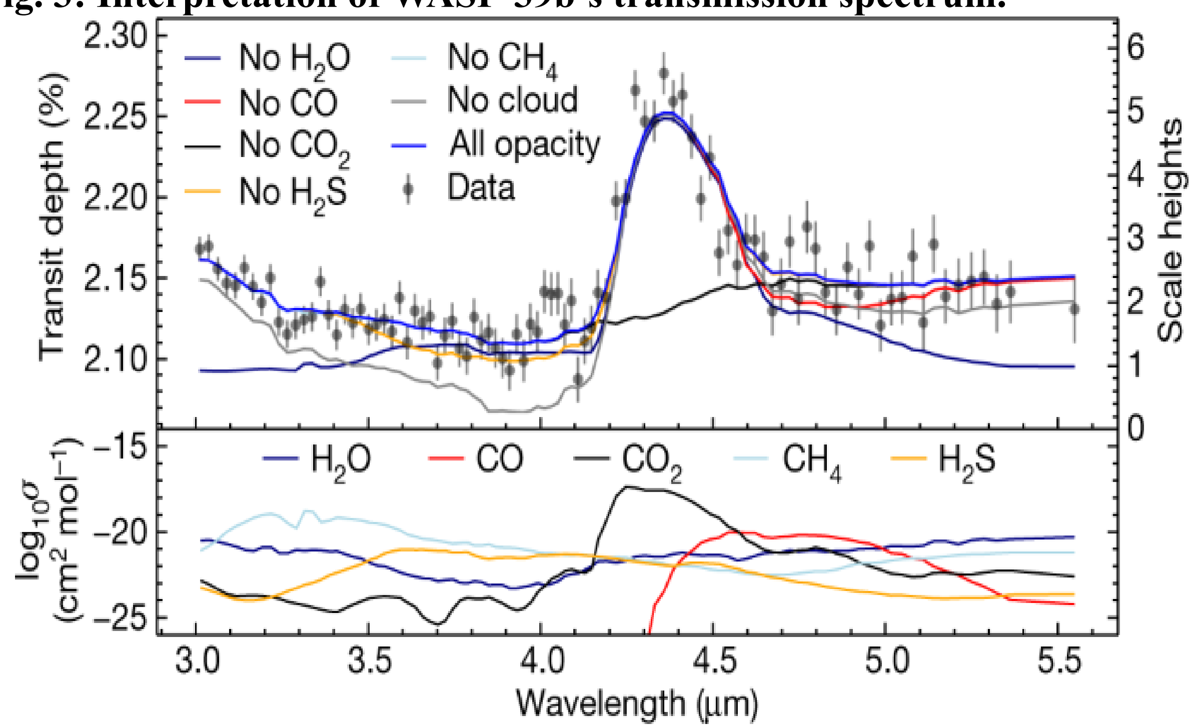


The JWST data (small coloured points) are compared with Spitzer's two Infrared Array Camera (IRAC) broadband photometric measurements (grey circles and corresponding sensitivity curves labelled IRAC1 and IRAC2). The axis on the right shows equivalent scale heights (750–1,000 km) in WASP-39b's atmosphere; for plotting purposes, we assume that one scale height corresponds to 800 km. The JWST data are consistent with the

Spitzer points (within 2σ) when integrated over the broad bandpasses (indicated by the horizontal lines). The relative transit depths between the 3.6-μm and 4.5-μm channels are also consistent within 2σ between independent reductions of the JWST data, with most of the deviation coming from the 3.6-μm bandpass. Vertical error bars indicate 1σ uncertainties.

We compared the data with a suite of one-dimensional atmospheric structure and transmission spectrum models to constrain the composition of WASP-39b's atmosphere. These models assume radiative–convective–thermochemical equilibrium, and they adopt a scaled solar abundance pattern. We calculated planet-specific grids of these models over a range of atmospheric metallicities, carbon-to-oxygen ratios and cloud properties using four different codes. These grids of self-consistent model transmission spectra were then fitted to the FIREFLy-reduced data (the fit results are independent of which dataset we use) while also adjusting for a reference radius at 1 bar. The results are illustrated in Fig. 3; see Methods for further details.

Fig. 3: Interpretation of WASP-39b's transmission spectrum.



Top: a comparison of the FIREFLy reduction and its 1σ uncertainties (labelled 'Data') to the best-fit ScCHIMERA theoretical model binned to the

resolution of the data (blue curve; Methods). The key parameters of the model are 10-times solar metallicity, a carbon-to-oxygen ratio of 0.35 and cloud opacity of 7×10^{-3} cm^2 g^{-1}. The impact of the opacity sources expected from thermochemical equilibrium over the full bandpass are indicated by removing the opacity contribution from individual gases one at a time. As in Fig. 2, the axis on the right shows equivalent scale heights in WASP-39b's atmosphere. Bottom: the molecular absorption cross-sections for each gas in the best-fit model. The model is well matched to the data ($X^2/N_{data} = 1.3$), suggesting that our assumptions broadly capture the important physics and chemistry in WASP-39b's atmosphere. However, there is a feature near 4.0 µm that cannot be reproduced by the models used here. The strong CO_2 absorption (4.1–4.6 µm) and the apparent lack of methane (3.0–3.5 µm) is what drives the solution to an elevated atmospheric metal enrichment, ruling out previous low-metallicity estimates[29,30,31]. The other reductions and models give similar results.

Under similar assumptions, all four model grids are able to match the dominant spectral morphologies—namely the strong CO_2 feature between 4.1 µm and 4.6 µm and the rise in transit depth bluewards of 3.6 µm owing to water (H_2O) vapour (a species that had been detected previously at shorter wavelengths[10]). More subtle modulations over the whole bandpass are potentially owing to contributions from clouds, carbon monoxide (CO) and hydrogen sulfide (H_2S), although the degree to which the two gas species contribute is unknown pending further study.

Several models for warm gas giant atmospheres predict that the CO_2 abundance scales quadratically with atmospheric metallicity, becoming detectable at 4.3 µm for metallicities above that of the Sun[1,2,3]. The representative best-fit model shown in Fig. 3 is consistent with this scenario. It has a 10-times-solar metal enrichment and a slightly subsolar carbon-to-oxygen ratio (0.35, compared with the solar value of 0.55; ref. [43]). The moderate contribution of cloud opacity predicted by the best-fit model is consistent with interpretations of previous population-level studies of planets that have similar temperatures and gravities to WASP-39b[44,45]. It is also consistent with the predictions of aerosol microphysics and global circulation models of hot giant planets[46,47,48].

In addition to the large CO_2 feature, we also identify a smaller spectral feature near 4.0 μm that is not matched by our thermochemical equilibrium models (Fig. 3). This feature is present in all four independent reductions and has a significance of 2σ (Methods). Further data analysis and modelling including non-equilibrium chemistry are needed to fully assess the robustness of this feature and to identify the chemical species that gives rise to it. Additional JWST ERS observations of WASP-39b that will use the G395H grating on NIRSpec also have the potential to confirm the 4.0-μm feature and resolve it in greater detail.

The grid fits explored here favour lower metallicities than refs. [10,21], and higher metallicities than ref. [31], even though the Spitzer data that their studies included are consistent with our JWST data. The higher precision and more resolved measurement of the CO_2 feature enabled by JWST pulls the models of refs. [10,21] to lower metallicity and increased cloudiness. Nevertheless, it is not possible to obtain a robust confidence interval on this inference without more rigorous Bayesian analyses, which is left to future work (Methods). Continued modelling of WASP-39b will also be aided by the future measurements of the planet's transmission spectrum from 0.5 μm to 5.5 μm that are also being obtained by this ERS programme. The final transmission spectrum will ultimately have higher spectral resolution than the data presented here (more than four times over most of the bandpass), and will be validated using multiple JWST instruments.

Methods

Data reduction

We reduced the JWST NIRSpec PRISM data for WASP-39b using four separate pipelines to confirm that the results did not depend on the specifics of the analyses, as was sometimes the case for results from the Spitzer Space Telescope (for example, ref. [49]). The descriptions below refer to calibration pipelines and other software whose code and citations appear in 'Code availability'.

tshirt pipeline

We used the Time Series Helper and Integration Reduction Tool[40] (tshirt) to extract light curves of the spectrum. This pipeline modifies the JWST Calibration pipeline steps to improve the precision of the reduction. tshirt has been used to successfully analyse the JWST transit observations of HAT-P-14b that were obtained during commissioning with the Near Infrared Camera (NIRCam)[37]. First, we used an updated bias frame from commissioning programme 1130 observation 29 and ran the JWST Calibration pipeline until the reference pixels step. We then applied a correction for $1/f$ noise (so named since the noise power is inversely proportional to the signal frequency, f), which varies for odd and even rows and for each column. We use background pixels for the calibration as reference pixels are not available in this subarray. We skipped the jump and dark subtraction steps because they were seen to add noise to the light curves. tshirt fits the profile of the spectrum with splines and rejects outlier pixels that are more than 50σ from the spline fits. We used covariance-weighted extraction[50] with an assumed pixel correlation of 0.08. For spectral extraction, we used a background region no closer than 7 pixels on either side of the source and an extraction region width of 16 pixels. The scatter in the light curve was consistent with the theoretical limit of photon and read noise over short timescales.

We fit the light curves with a second-order (quadratic) polynomial baseline, uninformative quadratic limb-darkening priors and an exponential start-up ramp with 10σ clipping of outliers. To begin, we fit the white-light curve with priors on the transit centre, inclination and period from ref. [22]. We also used the ratio a/R_* (where a is the semi-major axis and R_* is the stellar radius) from ref. [22] but widened the uncertainty on this parameter because the enforced prior resulted in significant residuals. Next, we fit each spectroscopic light curve individually with the orbital parameters fixed at the value from the white-light posterior medians. We modelled the light curves using the ‘exoplanet’ code[51] and the pymc3[52] sampler. We evaluated the wavelengths using the JWST Calibration pipeline at pixel row 16 ($Y = 16$) from the world coordinate solution. This uses an instrument model and could not be verified owing to a lack of strong stellar absorption features at the NIRSpec resolution. All the other reductions adopted this wavelength calibration. As shown in Fig. 1, the standard deviation in the out-of-transit light curve approaches the theoretical limit of photon and read noise at short

wavelengths, but is 20% to 40% higher at longer wavelengths, which may be related to uncorrected 1/*f* noise.

Eureka! pipeline

Eureka![39] is a data reduction and analysis pipeline for time-series observations with the JWST or the Hubble Space Telescope Its modular, multi-stage design provides flexibility and ease of comparison at any step, starting from uncalibrated FITS files and resulting in precise transmission or emission spectra. Eureka! has been used to successfully analyse the JWST transit observations of HAT-P-14b that were obtained during commissioning with NIRCam[37].

We began the data reduction process using the uncalibrated raw data files with the "uncal" suffix available from the Mikulski Archive for Space Telescopes (MAST) archive. The first stage of the Eureka! pipeline is primarily a wrapper for Stage 1 of the JWST Calibration pipeline, which converts groups to slopes. For this dataset, we skipped the jump detection step as it led to a large fraction of detector pixels being incorrectly flagged as outliers. We did, however, search for and flag outliers at multiple points in subsequent stages. We also manually updated the bad-pixel map to include identified hot pixels on the detector that were not provided in the current (July 2022) full-detector STScI data-quality map. As part of Eureka!, we performed a custom background subtraction at the group level before Stage 1 ramp fitting to account for 1/*f* noise introduced during detector readout. We set the top and bottom six rows of the detector as our background region and flagged pixels deemed outliers at $>3\sigma$. We then subtracted the mean flux per pixel column and repeated this for each group and integration in the observation. Similarly to Stage 1, the second stage of the Eureka! pipeline is a wrapper for Stage 2 of the JWST Calibration pipeline, which calibrates the two-dimensional time series of fitted slopes. Here, we skipped the flux calibration step, thus leaving the data in units of digital number (DN) per second (DN s^{-1}).

For Stage 3, we performed background subtraction and optimal extraction of the stellar spectrum for each integration with Eureka!. We used only pixels 14 to 495 in the dispersion direction of the 512 × 32-pixel subarray, as

NIRSpec's throughput is negligible beyond this range. We also masked pixels that have a non-zero data-quality flag to avoid any impact of outlier pixels on the extracted spectra or background subtraction. The position of the source on the detector along the cross-dispersion dimension is located by fitting a Gaussian to the pixel values summed over all detector columns. For each pixel, we examined its flux variation in time and performed a double-iteration, 10σ-outlier rejection test. We then executed a second column-by-column background subtraction, this time at the integration level, using pixels located at least 8 pixels away from the source position to compute the mean background per column. Performing this additional background subtraction reduced the number of outliers in the measured light curves and accounted for the residual background and/or noise introduced during the ramp fitting procedure. As with Stage 1, we exclude 3σ outliers from our background region. We adopted an aperture half-width of 7 pixels for our optimal spectral extraction step, constructing the profile from the median frame. At the end of this stage, we obtained a time series of one-dimensional spectra.

For the remaining stages, we used multiple pipelines (Eureka![39] and ExoTEP[53,54,55]) to generate and fit the light curves. We first generated median-normalized light curves at the instrument's native resolution (that is, from each detector column) using our Stage 3 outputs. We then clipped additional outliers in time for the white and spectroscopic light curves. For this step, we first rejected integrations that were more than 3σ outliers for the source position in the cross-dispersion direction, the width of the fitted Gaussian to the spatial profile or the drift in the dispersion direction. Next, we produced a median-filtered version of the light curve and clipped out 3σ outliers in flux. We jointly fit astrophysical and systematics model parameters to the white and individual spectroscopic light curves. Our astrophysical transit model used the batman package[56] with uniform priors, fitting for the following astrophysical parameters: the two coefficients of a stellar quadratic limb-darkening law, impact parameter, semi-major axis, transit time and the planet-to-star radius ratio (R_p/R_*) in each of the wavelength channels. Although the limb-darkening coefficients and planet-to-star radius ratio were fit independently in each spectroscopic channel, we used the best-fitting value of the planet's impact parameter, semi-major axis and transit time from a white-light curve fit as a fixed value in the

wavelength-dependent fits. For the systematics model, we assumed a linear trend in time for each wavelength channel, fitting for both the slope and *y* intercept. Last, we fit a single-point scatter to each light curve, which illustrates the level of additional noise required for our joint model to reach a reduced chi-squared (X^2) of unity. The white-light curve residuals have a root mean square (r.m.s.) of 3,013 ppm, and the spectroscopic light curves above 3 μm have a median r.m.s. of 5,779 ppm. Similar to the reduction shown in Fig. 1, both pipelines reach near photon noise. The Eureka! and ExoTEP transmission spectra appear nearly identical; therefore, only one (Eureka!) is shown in Fig. 2.

Tiberius pipeline

We built on the pipeline developed for the analysis of the Low Resolution Ground-Based Exoplanet Survey using Transmission Spectroscopy (LRG-BEASTS) data[21,57,58] to provide an independent reduction of the data. We began with the outputs of the JWST Calibration Stage 1 pipeline with the jump step correction turned off. We created bad-pixel and cosmic-ray masks by identifying 5σ outliers in running medians operating along pixel rows and along individual pixels in time. Before tracing the spectra, we interpolated each column of the detector onto a finer grid, ten times the initial spatial resolution, to improve the extraction of flux at the subpixel level. We used a fourth-order polynomial to trace the spectra and a four-pixel-wide aperture. To remove the 1/*f* noise, we fit a linear polynomial to 21 background pixels along each column in the cross-dispersion direction. Next, to correct for shifts in the dispersion direction, we cross-correlated each stellar spectrum with the first spectrum of the observation to account for very small (0.003–0.005) subpixel shifts. Our white-light curve spans a wavelength range of 0.518–5.348 μm after masking saturated pixels, and our 147 spectroscopic light curves used 3-pixel-wide bins across this same wavelength range. We masked frames 20751–20765 owing to a high-gain-antenna move that led to increased noise in the light curves.

We fit our light curves with a combination of a quadratically limb-darkened transit model (through batman[56]) with a linear-in-time polynomial. We began by fitting the white-light curve to derive the system parameters: inclination, *i*, time of mid-transit, T_C, the semi-major axis scaled to the

stellar radius, a/R_*, and the linear limb-darkening coefficient, u_1. We placed wide boundaries on the parameter values only to prevent unphysical values. In practice, the parameter values did not get close to the boundaries. We fixed the planet's orbital period to 4.0552941 d and the eccentricity to 0 from ref. [22]. We fixed the quadratic coefficient, u_2, to theoretical values determined by ExoTiC-LD[59,60] with three-dimensional stellar models[61], and fit for u_1. We used a Levenberg–Marquardt algorithm to fit our light curves, rescaled our photometric uncertainties to give a reduced $X^2 = 1$ for our best-fit model and then re-ran the fits. For the spectroscopic light curves, the system parameters (i, T_C and a/R_*) were held fixed to the best-fit values found from the white-light curve. The white-light curve residuals had an r.m.s. of 2,761 ppm and the spectroscopic light curve residuals had a median r.m.s. of 6,731 ppm. In both cases, the variance of the residuals scales upon binning as expected for Poisson noise.

FIREFLy pipeline

We also reduced the data using the FIREFLy reduction routines[38]. These routines utilize the JWST Calibration pipeline with custom modifications. This pipeline has been used to successfully analyse the JWST transit observations of HAT-P-14b that were obtained during commissioning with NIRSpec G395[37]. We removed $1/f$ noise (see ref. [36]) at the group level, as the $1/f$ noise changes from group to group. We also skipped the jump step and instead flagged and removed cosmic rays, bad pixels, hot pixels and other outliers using median filtering of the data both spatially and in time, flagging pixels using a 5σ-outlier threshold algorithm. The time series of two-dimensional spectra were aligned using cross-correlation and interpolation, with the time-series spectra exhibiting an r.m.s. jitter of 0.005 pixels in the x-axis direction and 0.0026 pixels in the y-axis direction. We found a small inverse ramp in the light curves, which settled down after the first 2,000 exposures, which we discarded. We fit the light curves with the batman[56] transit model along with a linear baseline and a second-order jitter detrending polynomial of x and y detector position as described by ref. [38], which are present in the spectrophotometry at the 53 ± 2-ppm level in the x direction and at the 140 ± 3-ppm level in the y direction. We applied a fixed quadratic limb-darkening law using the three-dimensional models[61]

computed using the methods of ref. [62] from ExoTiC-LD[59,60]. In fitting the 3 μm to 5.5 μm white-light curve, we allowed the semi-major axis in units of stellar radii a/R_*, inclination i and central transit time T_0 to freely vary along with the transit depth and systematics model. We used the Markov chain Monte Carlo sampling routine emcee[63] to find the best-fit parameters and measure the posterior distribution. We find the 3–5.5 μm white-light curve has a transit depth of 2.1368 ± 0.0014% and achieves 808-ppm scatter in the residuals. This is within 6% of the expected noise limit of 758 ppm as calculated by the JWST Calibration pipeline, with the scatter of the residuals decreasing to below 40 ppm upon binning with no detectable red noise. We fit each spectroscopic light curve shown in Fig. 1 with the same astrophysical and systematic models as the white-light curve, except fixing the system parameters (a/R_*, i and T_0). The transmission spectral light-curve residuals for each bin are typically within 5% of pipeline error or better, also with no detectable red noise.

Data–model comparison

We compared the extracted transmission spectral data to a suite of one-dimensional self-consistent radiative–convective–thermochemical equilibrium model atmospheres (see, for example, refs. [64,65] for a general description of such models) described below. In short, all models are able to fit the 3–5.5 μm spectra consistently (with $X^2/N_{data} < 1.4$, where N_{data} is the number of spectral data points) with a 10-times-solar metal enrichment and varying grey cloud opacity for their single best estimate. Comparisons of the model fits from each grid are shown in Extended Data Fig. 1. For additional parameters within the grid (for example, carbon-to-oxygen ratio (C/O) and heat redistribution), there is some discrepancy between each model grid's single best estimate values. Additional Bayesian analyses are needed to rigorously quantify confidence intervals on atmospheric properties of interest, which is beyond the scope of this work. Future works will focus on modelling that includes the effects of disequilibrium chemistry, aerosol microphysics and three-dimensional circulation effects. We assumed the following parameters in the modelling: stellar effective temperature, T_{eff} = 5,512 K, stellar radius = 0.932 $R_\odot$, planet mass = 0.281 M_J, planet radius = 1.279 R_J and planet orbital semi-major axis = 0.04828 au.

ScCHIMERA

This framework was first described in refs. [66,67], with the most recent updates, methods and opacity sources described in ref. [68]. We compute the converged atmospheric structure (temperature–pressure and thermochemical equilibrium gas mixing ratio profiles) over a grid of atmospheric metallicity ([M/H], where the square brackets indicate $\log_{10}$ enrichment relative to solar[43]) spaced at 0.25-dex intervals between 0 and 2.25 (1- to 175-times solar) and C/O at values of 0.20, 0.35, 0.55, 0.70, 0.75 and 0.80. We assume full day-to-night temperature redistribution[69] as planets in this temperature regime are unlikely to possess strong day-to-night temperature contrast[70,71]. We then compute transmission spectra[72,73] from these converged atmospheric structures. To match the models to the data, the DYNESTY[74] fitting routine is used to search for the optimal [M/H] and C/O (via nearest neighbour) while simultaneously adjusting the 1-bar planetary radius (which controls the absolute transit depth) and an opaque, grey, uniformly vertically distributed, cloud opacity ($\kappa_{\rm cld}$). The optimal model resulting from this process is $[M/H] = +1.0$, $C/O = 0.35$ and $\log_{10}\kappa_{\rm cld} = -2.15$ cm^2 g^{-1}. The metallicity and cloud opacity are primarily driven by the strength of the 4.3-µm CO_2 feature and lack of methane (CH_4) absorption near 3.3 µm. This result is what is shown in the main text (Fig. 3), which also illustrates the relative contribution of the key opacity sources (H_2O (refs. [75,76]), CO (refs. [77,78]), CO_2 (refs. [79,80]), H_2S (refs. [78,81]) and CH_4 (refs. [78,82])) to the overall spectral shape. Extended Data Fig. 2 shows the atmospheric structure (temperature profile and gas mixing ratio profiles) for this best-fit model.

PICASO

The core one-dimensional radiative–convective model is based on the legacy ‘Extrasolar Giant Planet’ code described in refs. [69,80,83] and since updated and modernized within the PICASO[84] framework described in ref. [85] (PICASO 3.0). The PICASO 3.0 model uses gaseous opacities created from the references listed in ref. [80]. The grid of PICASO models contains metallicity points at 0.1-, 0.3-, 1-, 3-, 10-, 30-, 50- and 100-times solar; C/O at 0.23, 0.46, 0.69 and 0.92; and also assumes full day–night heat

redistribution. The clouds are modelled using the Virga[86] implementation of the Eddysed[87] framework, which requires a vertical mixing coefficient, K_{zz} (constant with altitude; $\log_{10}K_{zz} = 5$, 7, 9 and 11 (cgs units)) and a vertically constant sedimentation parameter ($f_{sed} = 0.6$, 1, 3, 6 and 10), with optical/material properties for clouds thought to exist at WASP-39b's pressures and temperatures (Na_2S, MnS and $MgSiO_3$). The f_{sed} parameter controls the vertical extent of the cloud, and K_{zz} and f_{sed} together control the mean droplet sizes with altitude in the atmosphere. A X^2 grid search along the described dimensions is performed to identify the best fit. Within this grid, the nominal best fit ($X^2/N_{data} = 1.34$) is 10-times-solar metallicity, a subsolar C/O (0.23), with an extended large droplet cloud ($f_{sed} = 0.6$, $\log_{10}K_{zz} = 9$) that produces a grey continuum over these wavelengths, consistent with the ScCHIMERA results above.

ATMO

The ATMO radiative–convective–thermochemical equilibrium solver is described in refs. [88,89,90,91]. This grid consists of model transmission spectra for four different day–night energy redistribution factors (0.25, 0.5, 0.75 and 1.0, where 0.5 is 'full' and 1.0 is 'dayside only'), six metallicities (0.1-, 1-, 10-, 50-, 100- and 200-times solar), six C/O ratios (0.35, 0.55, 0.70, 0.75, 1.0 and 1.5), two haze factors (no haze and 10-times multi-gas Rayleigh scattering) and four grey cloud factors (no cloud, 0.5-, 1- and 5-times the strength of H_2 Rayleigh scattering at 350 nm between 1-mbar and 50-mbar pressure levels). Each model transmission spectrum from the grid is binned to the same resolution as that of the observations to compute X^2, with a (wavelength independent) transit depth offset as the free parameter. Within this grid, we find a best-fit model ($X^2/N_{data} = 1.39$) spectrum arising from a redistribution factor of 0.75 (slightly hotter than a full day–night redistribution would produce), a metallicity of 10-times solar, a super-solar C/O ratio of 0.7, a haze factor of 10 and a cloud factor of 5.

PHOENIX

This model originates from the PHOENIX stellar atmosphere code[92] adapted for exoplanets[93] with additional modelling and opacity updates described in refs. [94,95]. The model grid is computed for an array of irradiation temperatures (920 K, 1,020 K, 1,120 K and 1,220 K), metallicities (0.1-, 1-, 10- and 100-times solar) and C/O (0.3, 0.54, 0.7 and 1.0), and includes a sampling of opaque, grey clouds at specified cloud-top pressures. The nominal best-fit model ($X^2/N_{\mathrm{data}} = 1.32$) from this grid set-up results in a 10-times-solar metallicity and subsolar C/O (0.3) atmosphere with a cloud-top pressure of 0.3 mbar.

Quantifying feature detection significance

We quantified the detection significance[96] of CO_2 with the following steps. The best-fit grid model without CO_2 (that is, the 'no CO_2' black curve shown in Fig. 3) is first subtracted from the data, leaving behind a strong residual feature due to CO_2 (Extended Data Fig. 3). The peak per-spectral-bin mean signal-to-noise ratio of this residual feature is about 10σ. To utilize the full line/band shape we then fit the residual peak with (1) a four-parameter Gaussian model (centroid, amplitude, width and vertical offset), shown as red curves in Extended Data Fig. 3, and (2) a 'no feature' constant using a nested sampling routine[74]. The Bayesian evidence between the Gaussian model and constant model were then used to compute a Bayes factor, B, and corresponding detection significance[97]. For the CO_2 residual feature, $\ln(B)$ is 340.5, which equates to a 26.2σ detection. From this analysis, we conclude that the CO_2 feature is robustly detected.

On inspecting Figs. 2 and 3, there appears to be a feature near 4.0 µm (just short of the major CO_2 feature). We repeated the same analysis as above, but instead compared the Bayesian evidence from a two-component Gaussian model fit (to accommodate for both the CO_2 feature and the unknown absorber) to that of the single component Gaussian model fit above. On doing so, we find $\ln(B) = 0.98$, which equates to a 2σ significance. Restricting the prior range for the second Gaussian to be localized near the 4-µm feature boosts the significance to 2.3σ. Future analyses will focus on the nature of this feature and more rigorous quantification via nested

Bayesian model comparison within atmospheric retrieval frameworks (for example, ref. [34]).

Data availability

The data used in this paper are associated with JWST programme ERS 1366 (observation #4) and are available from the Mikulski Archive for Space Telescopes (https://mast.stsci.edu). Science data processing version (SDP_VER) 2022_2a generated the uncalibrated data that we downloaded from MAST. We used JWST calibration software version (CAL_VER) 1.5.3 with modifications described in the text. We used calibration reference data from context (CRDS_CTX) 0916, except as noted in the text. All the data and models presented in this publication can be found at https://doi.org/10.5281/zenodo.6959427. Source data are provided with this paper.

Code availability

The codes used in this publication to extract, reduce and analyse the data are as follows: STScI JWST Calibration pipeline[37] (https://github.com/spacetelescope/jwst), tshirt[40], Eureka![39] (https://eurekadocs.readthedocs.io/en/latest/), Tiberius[21,56,57] and FIREFLy[38]. In addition, these made use of Exoplanet[51] (https://docs.exoplanet.codes/en/latest/), Pymc3[52] (https://docs.pymc.io/en/v3/index.html), ExoTEP[53,54,55], Batman[56] (http://lkreidberg.github.io/batman/docs/html/index.html), ExoTiC-ISM[59] (https://github.com/Exo-TiC/ExoTiC-ISM), ExoTiC-LD[60] (https://exotic-ld.readthedocs.io/en/latest/), Emcee[63] (https://emcee.readthedocs.io/en/stable/), DYNESTY[74] (https://dynesty.readthedocs.io/en/stable/index.html) and chromatic (https://zkbt.github.io/chromatic/), each of which use the standard Python libraries scipy[98], numpy[99], astropy[100,101] and matplotlib[102]. The atmospheric models used to fit the data are as follows: PICASO[84] (https://natashabatalha.github.io/picaso/), Virga[86]

(https://natashabatalha.github.io/virga/), ScCHIMERA[68] (https://github.com/mrline/CHIMERA), ATMO[88,89,90,91] and PHOENIX[94].

References

1. Lodders, K. & Fegley, B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. I. Carbon, nitrogen, and oxygen. *Icarus* **155**, 393–424 (2002).

2. Zahnle, K., Marley, M. S., Freedman, R. S., Lodders, K. & Fortney, J. J. Atmospheric sulfur photochemistry on hot Jupiters. *Astrophys. J. Lett.* **701**, L20–L24 (2009).

3. Moses, J. I. et al. Compositional diversity in the atmospheres of hot Neptunes, with application to GJ 436b. *Astrophys. J.* **777**, 34 (2013).

4. Pollack, J. B. et al. Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).

5. Fortney, J. J. et al. A framework for characterizing the atmospheres of low-mass low-density transiting planets. *Astrophys. J.* **775**, 80 (2013).

6. Venturini, J., Alibert, Y. & Benz, W. Planet formation with envelope enrichment: new insights on planetary diversity. *Astron. Astrophys.* **596**, A90 (2016).

7. Kaltenegger, L. & Traub, W. A. Transits of Earth-like planets. *Astrophys. J.* **698**, 519–527 (2009).

8. Deming, D. et al. Discovery and characterization of transiting super Earths using an all-sky transit survey and follow-up by the James Webb Space Telescope. *Publ. Astron. Soc. Pac.* **121**, 952–967 (2009).

9. Lustig-Yaeger, J., Meadows, V. S. & Lincowski, A. P. The detectability and characterization of the TRAPPIST-1 exoplanet atmospheres with JWST. *Astron. J.* **158**, 27 (2019).

10. Wakeford, H. R. et al. The complete transmission spectrum of WASP-39b with a precise water constraint. *Astron. J.* **155**, 29 (2018).

11. Spake, J. J. et al. Abundance measurements of H_2O and carbon-bearing species in the atmosphere of WASP-127b confirm its supersolar metallicity. *Mon. Not. R. Astron. Soc.* **500**, 4042–4064 (2021).

12. Alderson, L. et al. A comprehensive analysis of WASP-17b's transmission spectrum from space-based observations. *Mon. Not. R. Astron. Soc.* **512**, 4185–4209 (2022).

13. Stevenson, K. B. et al. Transiting exoplanet studies and community targets for JWST's Early Release Science Program. *Publ. Astron. Soc. Pac.* **128**, 094401 (2016).

14. Bean, J. L. et al. The Transiting Exoplanet Community Early Release Science Program for JWST. *Publ. Astron. Soc. Pac.* **130**, 114402 (2018).

15. Faedi, F. et al. WASP-39b: a highly inflated Saturn-mass planet orbiting a late G-type star. *Astron. Astrophys.* **531**, A40 (2011).

16. Laughlin, G., Crismani, M. & Adams, F. C. On the anomalous radii of the transiting extrasolar planets. *Astrophys. J. Lett.* **729**, L7 (2011).

17. Thorngren, D. P. & Fortney, J. J. Bayesian analysis of hot-Jupiter radius anomalies: evidence for ohmic dissipation? *Astron. J.* **155**, 214 (2018).

18. Sarkis, P., Mordasini, C., Henning, T., Marleau, G. D. & Mollière, P. Evidence of three mechanisms explaining the radius anomaly of hot Jupiters. *Astron. Astrophys.* **645**, A79 (2021).

19. Fischer, P. D. et al. HST hot-Jupiter transmission spectral survey: clear skies for cool Saturn WASP-39b. *Astrophys. J.* **827**, 19 (2016).

20. Nikolov, N. et al. VLT FORS2 comparative transmission spectroscopy: detection of Na in the atmosphere of WASP-39b from the ground. *Astrophys. J.* **832**, 191 (2016).

21. Kirk, J. et al. LRG-BEASTS: transmission spectroscopy and retrieval analysis of the highly inflated Saturn-mass planet WASP-39b. *Astron. J.* **158**, 144 (2019).

22. Mancini, L. et al. The GAPS programme with HARPS-N at TNG. XVI. Measurement of the Rossiter–McLaughlin effect of transiting planetary systems HAT-P-3, HAT-P-12, HAT-P-22, WASP-39, and WASP-60. *Astron. Astrophys.* **613**, A41 (2018).

23. Biazzo, K. et al. The GAPS programme with HARPS-N at TNG. XXXV. Fundamental properties of transiting exoplanet host stars. *Astron. Astrophys.* **664**, A61 (2022).

24. Polanski, A. S., Crossfield, I. J. M., Howard, A. W., Isaacson, H. & Rice, M. Chemical abundances for 25 JWST exoplanet host stars with KeckSpec. *Res. Note. AAS* **6**, 155 (2022).

25. Lodders K. & Fegley B. *The Planetary Scientist's Companion* (Oxford Univ. Press, 1998).

26. Kreidberg, L. et al. A precise water abundance measurement for the hot Jupiter WASP-43b. *Astrophys. J. Lett.* **793**, L27 (2014).

27. Atreya, S. K. et al. The origin and evolution of Saturn: a post-Cassini perspective. Preprint at https://arxiv.org/abs/2205.06914 (2022).

28. Thorngren, D. & Fortney, J. J. Connecting giant planet atmosphere and interior modeling: constraints on atmospheric metal enrichment. *Astrophys. J. Lett.* **874**, L31 (2019).

29. Tsiaras, A. et al. A population study of gaseous exoplanets. *Astron. J.* **155**, 156 (2018).

30. Fisher, C. & Heng, K. Retrieval analysis of 38 WFC3 transmission spectra and resolution of the normalization degeneracy. *Mon. Not. R. Astron. Soc.* **481**, 4698–4727 (2018).

31. Pinhas, A., Madhusudhan, N., Gandhi, S. & MacDonald, R. H_2O abundances and cloud properties in ten hot giant exoplanets. *Mon. Not.*

R. Astron. Soc. **482**, 1485–1498 (2019).

32. Welbanks, L. et al. Mass–metallicity trends in transiting exoplanets from atmospheric abundances of H_2O, Na, and K. *Astrophys. J. Lett.* **887**, L20 (2019).

33. Min, M., Ormel, C. W., Chubb, K., Helling, C. & Kawashima, Y. The ARCiS framework for exoplanet atmospheres. Modelling philosophy and retrieval. *Astron. Astrophys.* **642**, A28 (2020).

34. Benneke, B. & Seager, S. How to distinguish between cloudy mini-Neptunes and water/volatile-dominated super-Earths. *Astrophys. J.* **778**, 153 (2013).

35. Jakobsen, P. et al. The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. I. Overview of the instrument and its capabilities. *Astron. Astrophys.* **661**, A80 (2022).

36. Birkmann, S. M. et al. The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. IV. Capabilities and predicted performance for exoplanet characterization. *Astron. Astrophys.* **661**, A83 (2022).

37. Rigby, J. et al. Characterization of JWST science performance from commissioning. Preprint at https://arxiv.org/abs/2207.05632 (2022).

38. Rustamkulov, Z., Sing, D. K., Liu, R. & Wang, A. Analysis of a JWST NIRSpec lab time series: characterizing systematics, recovering exoplanet transit spectroscopy, and constraining a noise floor. *Astrophys. J. Lett.* **928**, L7 (2022).

39. Bell, T. J. et al. Eureka!: an end-to-end pipeline for JWST time-series observations. *J. Open Source Sci.* **7**, 4503–4508 (2022).

40. Glidic, K. et al. Atmospheric characterization of hot Jupiter CoRoT-1 b using the Wide Field Camera 3 on the Hubble Space Telescope. *Astron. J.* **164**, 19 (2022).

41. Beichman, C. et al. Observations of transiting exoplanets with the James Webb Space Telescope (JWST). *Publ. Astron. Soc. Pac.* **126**, 1134–1173 (2014).

42. Yamamura, I., Tsuji, T. & Tanabé, T. AKARI observations of brown dwarfs. I. CO and CO_2 bands in the near-infrared spectra. *Astrophys. J.* **722**, 682–698 (2010).

43. Lodders, K., Palme, H. & Gail, H.-P. in *Solar System: Astronomy, Astrophysics, and Cosmology* Vol. 4B (ed. Trümper, J. E.) 712 (Landolt-Börnstein - Group VI ‘Astronomy and Astrophysics’, Springer, 2009).

44. Stevenson, K. B. Quantifying and predicting the presence of clouds in exoplanet atmospheres. *Astrophys. J. Lett.* **817**, L16 (2016).

45. Fu, G. et al. Statistical analysis of Hubble/WFC3 transit spectroscopy of extrasolar planets. *Astrophys. J. Lett.* **847**, L22 (2017).

46. Parmentier, V., Fortney, J. J., Showman, A. P., Morley, C. & Marley, M. S. Transitions in the cloud composition of hot Jupiters. *Astrophys. J.* **828**, 22 (2016).

47. Kataria, T. et al. The atmospheric circulation of a nine-hot-Jupiter sample: probing circulation and chemistry over a wide phase space. *Astrophys. J.* **821**, 9 (2016).

48. Gao, P. et al. Aerosol composition of hot giant exoplanets dominated by silicates and hydrocarbon hazes. *Nat. Astron.* **4**, 951–956 (2020).

49. Himes, M. D. & Harrington, J. On the dayside atmosphere of WASP-12b. *Astrophys. J.* **931**, 86 (2022).

50. Schlawin, E. et al. JWST noise floor. II. Systematic error sources in JWST NIRCam time series. *Astron. J.* **161**, 115 (2021).

51. Foreman-Mackey, D. et al. exoplanet: gradient-based probabilistic inference for exoplanet data & other astronomical time series. *J.Open Source Softw.* **6**, 3285 (2021).

52. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. PyMC3: python probabilistic programming framework. *Astrophysics Source Code Library* ascl:1610.016 (2016).

53. Benneke, B. et al. Spitzer observations confirm and rescue the habitable-zone super-Earth K2-18b for future characterization. *Astrophys. J.* **834**, 187 (2017).

54. Benneke, B. et al. Water vapor and clouds on the habitable-zone sub-Neptune exoplanet K2-18b. *Astrophys. J. Lett.* **887**, L14 (2019).

55. Benneke, B. et al. A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds. *Nat. Astron.* **3**, 813–821 (2019).

56. Kreidberg, L. batman: basic transit model calculation in Python. *Publ. Astron. Soc. Pac.* **127**, 1161–1165 (2015).

57. Kirk, J. et al. LRG-BEASTS III: ground-based transmission spectrum of the gas giant orbiting the cool dwarf WASP-80. *Mon. Not. R. Astron. Soc.* **474**, 876–885 (2018).

58. Kirk, J. et al. ACCESS and LRG-BEASTS: a precise new optical transmission spectrum of the ultrahot Jupiter WASP-103b. *Astron. J.* **162**, 34 (2021).

59. Laginja, I. & Wakeford, H. ExoTiC-ISM: a Python package for marginalised exoplanet transit parameters across a grid of systematic instrument models. *J Open Source Softw* **5**, 2281 (2020).

60. Wakeford, H. R. & Grant, D. Exo-TiC/ExoTiC-LD: ExoTiC-LD v2.1 Zenodo Release. *Zenodo* https://doi.org/10.5281/zenodo.6809899 (2022).

61. Magic, Z., Chiavassa, A., Collet, R. & Asplund, M. The Stagger-grid: a grid of 3D stellar atmosphere models. IV Limb darkening coefficients. *Astron. Astrophys.* **573**, A90 (2015).

62. Sing, D. K. Stellar limb-darkening coefficients for CoRot and Kepler. *Astron. Astrophys.* **510**, A21 (2010).

63. Foreman-Mackey, D. et al. emcee v3: a Python ensemble sampling toolkit for affine-invariant MCMC. *J. Open Source Softw.* **4**, 1864 (2019).

64. Marley, M. S. & Robinson, T. D. On the cool side: modeling the atmospheres of brown dwarfs and giant planets. *Ann. Rev. Astron. Astrophys.* **53**, 279–323 (2015).

65. Hubeny, I. Model atmospheres of sub-stellar mass objects. *Mon. Not. R. Astron. Soc.* **469**, 841–869 (2017).

66. Piskorz, D. et al. Ground- and space-based detection of the thermal emission spectrum of the transiting hot Jupiter KELT-2Ab. *Astron. J.* **156**, 133 (2018).

67. Arcangeli, J. et al. H^- opacity and water dissociation in the dayside atmosphere of the very hot gas giant WASP-18b. *Astrophys. J. Lett.* **855**, L30 (2018).

68. Mansfield, M. et al. A unique hot Jupiter spectral sequence with evidence for compositional diversity. *Nat. Astron.* **5**, 1224–1232 (2021).

69. Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D. & Freedman, R. Comparative planetary atmospheres: models of TrES-1 and HD 209458b. *Astrophys. J. Lett.* **627**, L69–L72 (2005).

70. Perez-Becker, D. & Showman, A. P. Atmospheric heat redistribution on hot Jupiters. *Astrophys. J.* **776**, 134 (2013).

71. Zhang, X. Atmospheric regimes and trends on exoplanets and brown dwarfs. *Res. Astron. Astrophys.* **20**, 099 (2020).

72. Mai, C. & Line, M. R. Exploring exoplanet cloud assumptions in JWST transmission spectra. *Astrophys. J.* **883**, 144 (2019).

73. Iyer, A. R. & Line, M. R. The influence of stellar contamination on the interpretation of near-infrared transmission spectra of sub-Neptune worlds around M-dwarfs. *Astrophys. J.* **889**, 78 (2020).

74. Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158 (2020).

75. Polyansky, O. L. et al. ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *Mon. Not. R. Astron. Soc.* **480**, 2597–2608 (2018).

76. Gharib-Nezhad, E. et al. EXOPLINES: molecular absorption cross-section database for brown dwarf and giant exoplanet atmospheres. *Astrophys. J. Suppl. Ser.* **254**, 34 (2021).

77. Li, G. et al. Rovibrational line lists for nine isotopologues of the CO molecule in the $X^1\Sigma^+$ ground electronic state. *Astrophys. J. Suppl. Ser.* **216**, 15 (2015).

78. Grimm, S. L. et al. HELIOS-K 2.0 opacity calculator and open-source opacity database for exoplanetary atmospheres. *Astrophys. J. Suppl. Ser.* **253**, 30 (2021).

79. Huang, X., Gamache, R. R., Freedman, R. S., Schwenke, D. W. & Lee, T. J. Reliable infrared line lists for 13 CO_2 isotopologues up to E'=18,000 cm^{-1} and 1500 K, with line shape parameters. *J. Quant. Spectrosc. Radiat. Transf.* **147**, 134–144 (2014).

80. Marley, M. S. et al. The Sonora brown dwarf atmosphere and evolution models. I. Model description and application to cloudless atmospheres in rainout chemical equilibrium. *Astrophys. J.* **920**, 85 (2021).

81. Azzam, A. A. A., Tennyson, J., Yurchenko, S. N. & Naumenko, O. V. ExoMol molecular line lists—XVI. The rotation–vibration spectrum of hot H_2S. *Mon. Not. R. Astron. Soc.* **460**, 4063–4074 (2016).

82. Hargreaves, R. J. et al. An accurate, extensive, and practical line list of methane for the HITEMP database. *Astrophys. J. Suppl. Ser.* **247**, 55 (2020).

83. Marley, M. S. & McKay, C. P. Thermal structure of Uranus' atmosphere. *Icarus* **138**, 268–286 (1999).

84. Batalha, N. E., Marley, M. S., Lewis, N. K. & Fortney, J. J. Exoplanet reflected-light spectroscopy with PICASO. *Astrophys. J.* **878**, 70 (2019).

85. Mukherjee, S., Batalha, N. E., Fortney, J. J. & Marley, M. S. PICASO 3.0: a one-dimensional climate model for giant planets and brown dwarfs. Preprint at https://arXiv.org/abs/2208.07836 (2022).

86. Rooney, C. M., Batalha, N. E., Gao, P. & Marley, M. S. A new sedimentation model for greater cloud diversity in giant exoplanets and brown dwarfs. *Astrophys. J.* **925**, 33 (2022).

87. Ackerman, A. S. & Marley, M. S. Precipitating condensation clouds in substellar atmospheres. *Astrophys. J.* **556**, 872 (2001).

88. Tremblin, P. et al. Fingering convection and cloudless models for cool brown dwarf atmospheres. *Astrophys. J.* **804**, L17 (2015).

89. Drummond, B. et al. The effects of consistent chemical kinetics calculations on the pressure–temperature profiles and emission spectra of hot Jupiters. *Astron. Astrophys.* **594**, A69 (2016).

90. Goyal, J. M. et al. A library of ATMO forward model transmission spectra for hot Jupiter exoplanets. *Mon. Not. R. Astron. Soc.* **474**, 5158–5185 (2018).

91. Goyal, J. M. et al. A library of self-consistent simulated exoplanet atmospheres. *Mon. Not. R. Astron. Soc.* **498**, 4680–4704 (2020).

92. Hauschildt, P. H., Allard, F. & Baron, E. The NextGen Model atmosphere grid for $3000 \leq T_{\mathrm{eff}} \leq 10{,}000$ K. *Astrophys. J.* **512**, 377 (1999).

93. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885 (2001).

94. Lothringer, J. D. & Barman, T. The PHOENIX exoplanet retrieval algorithm and using H^- opacity as a probe in ultrahot Jupiters. *Astron. J.* **159**, 289 (2020).

95. Lothringer, J. D. et al. New window into planet formation and migration: refractory-to-volatile elemental ratios in ultra-hot Jupiters. *Astrophys. J.* **914**, 12 (2021).

96. Gregory, P. C. (ed.) *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support* (Cambridge Univ. Pres, 2005).

97. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104 (2008).

98. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

99. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

100. The Astropy Collaboration Astropy: a community Python package for astronomy. *Astron. Astrophys.* **558**, A33 (2013).

101. The Astropy Collaboration The Astropy Project: building an open-science project and status of the v2.0 core package. *Astron. J.* **156**, 123 (2018).

102. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

Acknowledgements

This work is based on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive

for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. These observations are associated with programme JWST-ERS-01366. Support for programme JWST-ERS-01366 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127. The results reported herein benefited during the design phase from collaborations and/or information exchange within NASA's Nexus for Exoplanet System Science (NExSS) research coordination network sponsored by NASA's Science Mission Directorate. N.M.B. acknowledges funding from NASA's Interdisciplinary Consortia for Astrobiology Research (NNH19ZDA001N-ICAR) under award number 19-ICAR19_2-0041.

Author information

Authors and Affiliations

1. Department of Physics, University of Warwick, Coventry, UK

   Eva-Maria Ahrer & Peter J. Wheatley

2. Centre for Exoplanets and Habitability, University of Warwick, Coventry, UK

   Eva-Maria Ahrer & Peter J. Wheatley

3. School of Physics, University of Bristol, Bristol, UK

   Lili Alderson, Hannah R. Wakeford & David Grant

4. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

   Natalie M. Batalha, Aarynn L. Carter, Jonathan J. Fortney, Sagnick Mukherjee, Kazumasa Ohno & Nicholas Scarsdale

5. NASA Ames Research Center, Moffett Field, CA, USA

Natasha E. Batalha

6. Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA

Jacob L. Bean & Adina D. Feinstein

7. Department of Astronomy, University of Wisconsin-Madison, Madison, WI, USA

Thomas G. Beatty

8. BAER Institute, NASA Ames Research Center, Moffett Field, CA, USA

Taylor J. Bell

9. Department of Physics and Institute for Research on Exoplanets, Université de Montréal, Montreal, Quebec, Canada

Björn Benneke, Caroline Piaulet, Michael Radica, Pierre-Alexis Roy, Jake Taylor & Louis-Philippe Coulombe

10. Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, USA

Zachory K. Berta-Thompson, Catriona Anne Murray, Molly Nehring, William C. Waalkes & Patcharapol Wachiraphan

11. Department of Physics and Astronomy, University of Kansas, Lawrence, KS, USA

Ian J. M. Crossfield & Jonathan Brande

12. Space Telescope Science Institute, Baltimore, MD, USA

Néstor Espinoza, Nikolay K. Nikolov, Leonardo A. Dos Santos & Jeff A. Valenti

13. Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA

Néstor Espinoza, David K. Sing, Natalie H. Allen, Guangwei Fu & Lakeisha Ramos-Rosado

14. School of Physics, Trinity College Dublin, Dublin, Ireland

Neale P. Gibson

15. School of Earth and Planetary Sciences (SEPS), National Institute of Science Education and Research (NISER), HBNI, Jatani, India

Jayesh M. Goyal

16. Department of Astronomy, University of Maryland, College Park, MD, USA

Eliza M.-R. Kempton, Guangwei Fu & Thaddeus D. Komacek

17. Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA

James Kirk, Mercedes López-Morales, Kevin Ortiz Ceballos & Diana Powell

18. Max Planck Institute for Astronomy, Heidelberg, Germany

Laura Kreidberg, Karan Molaverdikhani, Thomas Henning, Luigi Mancini, Thomas Mikal-Evans, Maria E. Steinrueck & Sebastian Zieba

19. School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

Michael R. Line, Luis Welbanks & Evgenya L. Shkolnik

20. Department of Physics, Utah Valley University, Orem, UT, USA

Joshua D. Lothringer

21. Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA

Sarah E. Moran & Mark S. Marley

22. Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

Vivien Parmentier, Xianyu Tan & Shang-Min Tsai

23. Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Nice, France

Vivien Parmentier

24. Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA

Zafar Rustamkulov & David K. Sing

25. Steward Observatory, University of Arizona, Tucson, AZ, USA

Everett Schlawin & Megan Mansfield

26. Johns Hopkins APL, Laurel, MD, USA

Kevin B. Stevenson, E. M. May & Jacob Lustig-Yaeger

27. European Space Agency, Space Telescope Science Institute, Baltimore, MD, USA

Stephan M. Birkmann & Sarah Kendrew

28. Leiden Observatory, University of Leiden, Leiden, The Netherlands

Nicolas Crouzet, Yamila Miguel & Sebastian Zieba

29. INAF – Osservatorio Astrofisico di Torino, Turin, Italy

Patricio E. Cubillos & Luigi Mancini

30. Space Research Institute, Austrian Academy of Sciences, Graz, Austria

Patricio E. Cubillos

31. Astrophysics Section, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Mario Damiano, Renyu Hu, Tiffany Kataria & Gael M. Roudier

32. Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, The Netherlands

Jean-Michel Désert

33. Earth and Planets Laboratory, Carnegie Institution for Science, Washington DC, USA

Peter Gao, Munazza K. Alam, Anjali A. A. Piette & Johanna K. Teske

34. Planetary Sciences Group, Department of Physics and Florida Space Institute, University of Central Florida, Orlando, FL, USA

Joseph Harrington

35. Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

Renyu Hu, Heather A. Knutson, Julie Inglis, Jessica J. Spake & Nicole L. Wallack

36. Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, Gif-sur-Yvette, France

Pierre-Olivier Lagage & A. García Muñoz

37. Laboratoire d'Astrophysique de Bordeaux, Université de Bordeaux, Pessac, France

Jérémy Leconte

38. Département d’Astronomie, Université de Genève, Sauverny, Switzerland

Monika Lendl & D. J. M. Petit dit de la Roche

39. Department of Astronomy and Carl Sagan Institute, Cornell University, Ithaca, NY, USA

Ryan J. MacDonald, Laura Flagg, Yu-Cian Hong, Nikole K. Lewis & Jake D. Turner

40. SRON Netherlands Institute for Space Research, Leiden, The Netherlands

Yamila Miguel

41. Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Munich, Germany

Karan Molaverdikhani

42. Exzellenzcluster Origins, Garching, Germany

Karan Molaverdikhani

43. Space Science Institute, Boulder, CO, USA

Julianne I. Moses

44. Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA

Keivan G. Stassun

45. Indian Institute of Technology, Indore, India

Keshav Aggarwal

46. School of Physical Sciences, The Open University, Milton Keynes, UK

Agnibha Banerjee & Joanna K. Barstow

47. Department of Physics, New York University Abu Dhabi, Abu Dhabi, UAE

Jasmina Blecic

48. School of Physics and Astronomy, University of Leicester, Leicester, UK

S. L. Casewell & Michael T. Roman

49. Department of Physics and Astronomy, University College London, London, UK

Quentin Changeat & Ingo P. Waldmann

50. Centre for Exoplanet Science, University of St Andrews, St Andrews, UK

K. L. Chubb

51. NASA Goddard Space Flight Center, Greenbelt, MD, USA

Knicole D. Colón & Avi M. Mandell

52. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

Tansu Daylan

53. Planetary Science Institute, Tucson, AZ, USA

Miguel de Val-Borro

54. Institute of Astronomy, Department of Physics and Astronomy, KU Leuven, Leuven, Belgium

Leen Decin

55. Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO, USA

Kevin France

56. Department of Physics and Astronomy, University of Delaware, Newark, DE, USA

John E. Gizis

57. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Ana Glidden & Benjamin V. Rackham

58. Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Ana Glidden & Benjamin V. Rackham

59. University Observatory Munich, Ludwig Maximilian University, Munich, Germany

Kevin Heng

60. Institute for Astrophysics, University of Vienna, Vienna, Austria

Nicolas Iro

61. California Institute of Technology, IPAC, Pasadena, CA, USA

Jessica E. Krick

62. Center for Space and Habitability, University of Bern, Bern, Switzerland

Elspeth K. H. Lee

63. Departamento de Astrofísica, Centro de Astrobiología (CAB, CSIC-INTA), Madrid, Spain

Jorge Lillo-Box

64. Department of Physics, University of Rome “Tor Vergata”, Rome, Italy

Luigi Mancini

65. Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain

Giuseppe Morello

66. Departamento de Astrofísica, Universidad de La Laguna (ULL), Tenerife, Spain

Giuseppe Morello

67. INAF- Palermo Astronomical Observatory, Palermo, Italy

Giuseppe Morello

68. Institute of Astronomy, University of Cambridge, Cambridge, UK

Matthew C. Nixon & Laura K. Rogers

69. Department of Astronomy, University of Michigan, Ann Arbor, MI, USA

Emily Rauscher

70. Astronomy Department and Van Vleck Observatory, Wesleyan University, Middletown, CT, USA

Seth Redfield

71. Universidad Adolfo Ibáñez: Penalolen, Santiago, Chile

Michael T. Roman

72. Astrophysics Group, Keele University, Staffordshire, UK

John Southworth

73. Maison de la Simulation, CEA, CNRS, Université Paris-Sud, UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

Pascal Tremblin

74. Department of Physics, Brown University, Providence, RI, USA

Gregory S. Tucker

75. Université de Paris Cité and Université Paris Est Creteil, CNRS, LISA, Paris, France

Olivia Venot

76. Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA

Xi Zhang

Consortia

JWST Transiting Exoplanet Community Early Release Science Team

- Eva-Maria Ahrer
- , Lili Alderson
- , Natalie M. Batalha
- , Natasha E. Batalha
- , Jacob L. Bean
- , Thomas G. Beatty
- , Taylor J. Bell
- , Björn Benneke
- , Zachory K. Berta-Thompson
- , Aarynn L. Carter
- , Ian J. M. Crossfield
- , Néstor Espinoza
- , Adina D. Feinstein
- , Jonathan J. Fortney

- , Neale P. Gibson
- , Jayesh M. Goyal
- , Eliza M.-R. Kempton
- , James Kirk
- , Laura Kreidberg
- , Mercedes López-Morales
- , Michael R. Line
- , Joshua D. Lothringer
- , Sarah E. Moran
- , Sagnick Mukherjee
- , Kazumasa Ohno
- , Vivien Parmentier
- , Caroline Piaulet
- , Zafar Rustamkulov
- , Everett Schlawin
- , David K. Sing
- , Kevin B. Stevenson
- , Hannah R. Wakeford
- , Natalie H. Allen
- , Stephan M. Birkmann
- , Jonathan Brande
- , Nicolas Crouzet
- , Patricio E. Cubillos
- , Mario Damiano
- , Jean-Michel Désert
- , Peter Gao
- , Joseph Harrington
- , Renyu Hu
- , Sarah Kendrew
- , Heather A. Knutson
- , Pierre-Olivier Lagage
- , Jérémy Leconte
- , Monika Lendl
- , Ryan J. MacDonald
- , E. M. May
- , Yamila Miguel
- , Karan Molaverdikhani

- , Julianne I. Moses
- , Catriona Anne Murray
- , Molly Nehring
- , Nikolay K. Nikolov
- , D. J. M. Petit dit de la Roche
- , Michael Radica
- , Pierre-Alexis Roy
- , Keivan G. Stassun
- , Jake Taylor
- , William C. Waalkes
- , Patcharapol Wachiraphan
- , Luis Welbanks
- , Peter J. Wheatley
- , Keshav Aggarwal
- , Munazza K. Alam
- , Agnibha Banerjee
- , Joanna K. Barstow
- , Jasmina Blecic
- , S. L. Casewell
- , Quentin Changeat
- , K. L. Chubb
- , Knicole D. Colón
- , Louis-Philippe Coulombe
- , Tansu Daylan
- , Miguel de Val-Borro
- , Leen Decin
- , Leonardo A. Dos Santos
- , Laura Flagg
- , Kevin France
- , Guangwei Fu
- , A. García Muñoz
- , John E. Gizis
- , Ana Glidden
- , David Grant
- , Kevin Heng
- , Thomas Henning
- , Yu-Cian Hong

- , Julie Inglis
- , Nicolas Iro
- , Tiffany Kataria
- , Thaddeus D. Komacek
- , Jessica E. Krick
- , Elspeth K. H. Lee
- , Nikole K. Lewis
- , Jorge Lillo-Box
- , Jacob Lustig-Yaeger
- , Luigi Mancini
- , Avi M. Mandell
- , Megan Mansfield
- , Mark S. Marley
- , Thomas Mikal-Evans
- , Giuseppe Morello
- , Matthew C. Nixon
- , Kevin Ortiz Ceballos
- , Anjali A. A. Piette
- , Diana Powell
- , Benjamin V. Rackham
- , Lakeisha Ramos-Rosado
- , Emily Rauscher
- , Seth Redfield
- , Laura K. Rogers
- , Michael T. Roman
- , Gael M. Roudier
- , Nicholas Scarsdale
- , Evgenya L. Shkolnik
- , John Southworth
- , Jessica J. Spake
- , Maria E. Steinrueck
- , Xianyu Tan
- , Johanna K. Teske
- , Pascal Tremblin
- , Shang-Min Tsai
- , Gregory S. Tucker
- , Jake D. Turner

- , Jeff A. Valenti
- , Olivia Venot
- , Ingo P. Waldmann
- , Nicole L. Wallack
- , Xi Zhang
- & Sebastian Zieba

Contributions

N.M.B., J.L.B. and K.B.S. provided overall programme leadership and management. D.K.S., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., K.B.S., L.K., M.L.-M., M.R.L., N.M.B., V.P. and Z.K.B.-T. made significant contributions to the design of the programme. K.B.S. generated the observing plan with input from the team. E.S., N.E. and T.G.B. provided instrument expertise. B.B., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., L.K., M.L.-M., M.R.L., N.M.B. and Z.K.B.-T. led or co-led working groups and/or contributed to significant strategic planning efforts like the design and implementation of the pre-launch Data Challenges. A.L.C., D.K.S., E.S., N.E., N.P.G., T.G.B. and V.P. generated simulated data for pre-launch testing of methods. A.L.C., A.D.F., C.P., E.-M.A., E.S., J.K., L.A., T.J.B. and Z.R. contributed to the development of data analysis pipelines and/or provided the data analysis products used in this analysis, reduced the data, modelled the light curves, and produced the planetary spectrum. J.M.G., J.D.L., K.O., M.R.L., N.E.B., S.M. and S.E.M. generated theoretical model grids for comparison with data. A.L.C., D.K.S., E.S., H.R.W., J.K., J.J.F., J.M.G., J.L.B., J.D.L., K.B.S., K.O., M.R.L., N.E.B., N.M.B., S.M., S.E.M., Z.K.B.-T. and Z.R. contributed significantly to the writing of this manuscript. E.-M.A., E.S., M.R.L. and Z.K.B.-T. generated figures for this manuscript. Over 100 of the co-authors participated in the design of this programme and contributed to the ERS observing proposal. Nearly all co-authors participated in pre-launch working groups and programme events. Nearly all authors provided scientific and technical input on the manuscript.

Corresponding author

Correspondence to Natalie M. Batalha.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Masahiro Ikoma and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Comparison of transmission spectrum modelling results from different codes for WASP-39b.

Despite different radiative–convective equilibrium and chemical solvers, treatments of clouds, grid spacing and grid-fitting approaches, all four grids arrive at the same 10-times solar metallicity point solution. Additionally, all four provide an acceptable fit to the data, with best-fitting $X^2/N_{data} < 1.4$.

Extended Data Fig. 2 Atmospheric structure arising from the best-fit model.

The thick red curve (and corresponding top *x* axis) shows the resulting 1D radiative–convective equilibrium temperature profile. The dashed lines (and bottom *x* axis) show the vertical gas mixing ratio profiles under the

assumption of thermochemical equilibrium. These abundances, along with the absorption cross-sections shown in the bottom panel of Fig. 3, are what control the relative contributions of each gaseous opacity to the total transmission spectrum.

Extended Data Fig. 3 Assessment of the strength of spectral features for WASP-39b.

Residual features (blue data points) after subtracting the continuum best model (black ‘no CO_2’ model curve in Fig. 3). A best-fitting ensemble of a two-component Gaussian model to both the CO_2 feature and the unknown absorber feature (~4 μm) is shown in red.

Supplementary information

Supplementary Data

This file contains files to recreate the main text figures. It includes: TRANSMISSION_SPECTRA_DATA, containing the data from each reduction; MODEL_FITS, containing the best-fitting model spectra from each of the four model grids described in the manuscript as well as the ‘remove one gas at a time’ spectra to recreate the top panel of Fig. 3; and the data necessary to reproduce the light curves shown in Fig. 1 are contained in the CSV file.

Source data

Source Data Fig. 1

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to

the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

JWST Transiting Exoplanet Community Early Release Science Team. Identification of carbon dioxide in an exoplanet atmosphere. *Nature* **614**, 649–652 (2023). https://doi.org/10.1038/s41586-022-05269-w

- Received: 04 August 2022
- Accepted: 23 August 2022
- Published: 02 September 2022
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05269-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM**
    - Z. Rustamkulov
    - D. K. Sing
    - S. Zieba

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRISS**
    - Adina D. Feinstein
    - Michael Radica
    - Xi Zhang

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H**
    - Lili Alderson
    - Hannah R. Wakeford
    - Xi Zhang

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRCam**
    - Eva-Maria Ahrer
    - Kevin B. Stevenson

- Xi Zhang

Nature (2023)

Atmospheric abundance

- Stefanie Reichert

Nature Physics News & Views 06 Oct 2022

JWST opens a window on exoplanet skies

- Julia V. Seidel
- Louise D. Nielsen
- Subhajit Sarkar

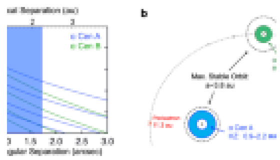
Nature News & Views Forum 15 Feb 2023

A population of red candidate massive galaxies ~600 Myr after the Big Bang

- Ivo Labbé
- Pieter van Dokkum
- Bingjie Wang

Nature Article 22 Feb 2023

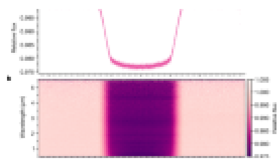
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05269-w



Imaging low-mass planets within the habitable zone of α Centauri

10 February 2021

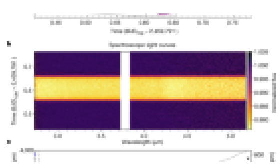
K. Wagner, A. Boehle, … T. de Zeeuw



Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM

09 January 2023

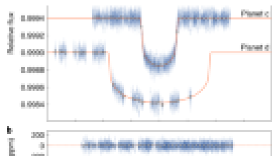
Z. Rustamkulov, D. K. Sing, … S. Zieba



Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H

09 January 2023

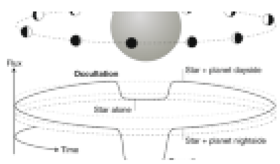
Lili Alderson, Hannah R. Wakeford, … Xi Zhang



[Transit detection of the long-period volatile-rich super-Earth ν2 Lupi d with CHEOPS]

28 June 2021

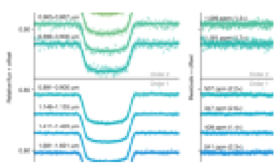
Laetitia Delrez, David Ehrenreich, … Nicholas A. Walton



[Highlights of exoplanetary science from Spitzer]

14 May 2020

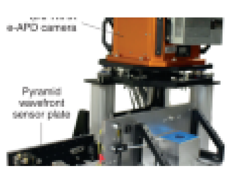
Drake Deming & Heather A. Knutson



[Early Release Science of the exoplanet WASP-39b with JWST NIRISS]

09 January 2023

Adina D. Feinstein, Michael Radica, … Xi Zhang



[Enabling high-res exoplanet spectroscopy at Keck]

16 July 2021

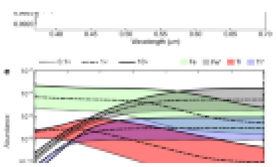
Dimitri Mawet



[The Habitable Exoplanet Observatory]

01 August 2018

B. Scott Gaudi, Sara Seager, … Habitable Exoplanet Observatory Science and Technology Definition Team



[Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b]

15 August 2018

H. Jens Hoeijmakers, David Ehrenreich, … Luca Di Fabrizio

- Article
- Open Access
- Published: 09 January 2023

Early Release Science of the exoplanet WASP-39b with JWST

NIRCam

- Eva-Maria Ahrer ,
- Kevin B. Stevenson ,
- Megan Mansfield,
- Sarah E. Moran,
- Jonathan Brande,
- Giuseppe Morello,
- Catriona A. Murray,
- Nikolay K. Nikolov,
- Dominique J. M. Petit dit de la Roche,
- Everett Schlawin,
- Peter J. Wheatley,
- Sebastian Zieba,
- Natasha E. Batalha,
- Mario Damiano,
- Jayesh M. Goyal,
- Monika Lendl,
- Joshua D. Lothringer,
- Sagnick Mukherjee,
- Kazumasa Ohno,
- Natalie M. Batalha,
- Matthew P. Battley,
- Jacob L. Bean,
- Thomas G. Beatty,
- Björn Benneke,
- Zachory K. Berta-Thompson,
- Aarynn L. Carter,
- Patricio E. Cubillos,
- Tansu Daylan,
- Néstor Espinoza,
- Peter Gao,
- Neale P. Gibson,
- Samuel Gill,
- Joseph Harrington,
- Renyu Hu,

- Laura Kreidberg,
- Nikole K. Lewis,
- Michael R. Line,
- Mercedes López-Morales,
- Vivien Parmentier,
- Diana K. Powell,
- David K. Sing,
- Shang-Min Tsai,
- Hannah R. Wakeford,
- Luis Welbanks,
- Munazza K. Alam,
- Lili Alderson,
- Natalie H. Allen,
- David R. Anderson,
- Joanna K. Barstow,
- Daniel Bayliss,
- Taylor J. Bell,
- Jasmina Blecic,
- Edward M. Bryant,
- Matthew R. Burleigh,
- Ludmila Carone,
- S. L. Casewell,
- Quentin Changeat,
- Katy L. Chubb,
- Ian J. M. Crossfield,
- Nicolas Crouzet,
- Leen Decin,
- Jean-Michel Désert,
- Adina D. Feinstein,
- Laura Flagg,
- Jonathan J. Fortney,
- John E. Gizis,
- Kevin Heng,
- Nicolas Iro,
- Eliza M.-R. Kempton,
- Sarah Kendrew,
- James Kirk,

- Heather A. Knutson,
- Thaddeus D. Komacek,
- Pierre-Olivier Lagage,
- Jérémy Leconte,
- Jacob Lustig-Yaeger,
- Ryan J. MacDonald,
- Luigi Mancini,
- E. M. May,
- N. J. Mayne,
- Yamila Miguel,
- Thomas Mikal-Evans,
- Karan Molaverdikhani,
- Enric Palle,
- Caroline Piaulet,
- Benjamin V. Rackham,
- Seth Redfield,
- Laura K. Rogers,
- Pierre-Alexis Roy,
- Zafar Rustamkulov,
- Evgenya L. Shkolnik,
- Kristin S. Sotzen,
- Jake Taylor,
- P. Tremblin,
- Gregory S. Tucker,
- Jake D. Turner,
- Miguel de Val-Borro,
- Olivia Venot &
- …
- Xi Zhang

Show authors

Nature **volume 614**, pages 653–658 (2023)

- 4297 Accesses

- 3 Citations

- 18 Altmetric
- Metrics details

Abstract

Measuring the metallicity and carbon-to-oxygen (C/O) ratio in exoplanet atmospheres is a fundamental step towards constraining the dominant chemical processes at work and, if in equilibrium, revealing planet formation histories. Transmission spectroscopy (for example, refs. [1,2]) provides the necessary means by constraining the abundances of oxygen- and carbon-bearing species; however, this requires broad wavelength coverage, moderate spectral resolution and high precision, which, together, are not achievable with previous observatories. Now that JWST has commenced science operations, we are able to observe exoplanets at previously uncharted wavelengths and spectral resolutions. Here we report time-series observations of the transiting exoplanet WASP-39b using JWST's Near InfraRed Camera (NIRCam). The long-wavelength spectroscopic and short-wavelength photometric light curves span 2.0–4.0 micrometres, exhibit minimal systematics and reveal well defined molecular absorption features in the planet's spectrum. Specifically, we detect gaseous water in the atmosphere and place an upper limit on the abundance of methane. The otherwise prominent carbon dioxide feature at 2.8 micrometres is largely masked by water. The best-fit chemical equilibrium models favour an atmospheric metallicity of 1–100-times solar (that is, an enrichment of elements heavier than helium relative to the Sun) and a substellar C/O ratio. The inferred high metallicity and low C/O ratio may indicate significant accretion of solid materials during planet formation (for example, refs. [3,4,]) or disequilibrium processes in the upper atmosphere (for example, refs. [5,6]).

Main

JWST has demonstrated the necessary precision and wavelength coverage to make bulk characterization of hot exoplanet atmospheres routine[7]. The JWST director's discretionary Early Release Science (ERS) programme provides the scientific community with observations of typical targets

quickly enough to inform planning for the telescope's second cycle of scheduled observations. The primary goals of the Transiting Exoplanet Community ERS programme (ERS 1366, led by N. M. Batalha, J. L. Bean and K. B. Stevenson) are to demonstrate instrument capabilities, quickly build community experience and seed initial discovery in transiting exoplanetary science[8,9]. The Panchromatic Transmission programme observed a single exoplanet, WASP-39b, in transmission using four different instrument modes. It included overlapping wavelength coverage to cross-compare and validate all three near-infrared instruments for time-series observations. The observations presented here form one-quarter of this programme, demonstrating the capacity of the JWST Near-InfraRed Camera (NIRCam) for transiting exoplanet atmospheric characterization.

WASP-39b is a highly inflated exoplanet of roughly Saturn mass, orbiting its G7 main-sequence star with a 4.05-day period[10]. We selected WASP-39b for its inactive host star and prominent spectroscopic features, which trace the atmospheric composition of the planet. We confirmed the star's relative inactivity through a photometric monitoring campaign using the Next-Generation Transit Survey (NGTS)[11] and Transiting Exoplanet Survey Satellite (TESS)[12] (Methods). Reported atmospheric metallicities span a range of possible values (0.003–300× solar)[13,14,15,16,17,18] owing to limits on wavelength coverage, lower signal-to-noise ratio data and/or differences between analyses[19,20,21,22]. If the Solar System trend for gas giants[23,24] also applies to exoplanets, WASP-39b should have an atmospheric metallicity comparable to that of Saturn (10× solar[25]) and other Saturn-mass exoplanets.

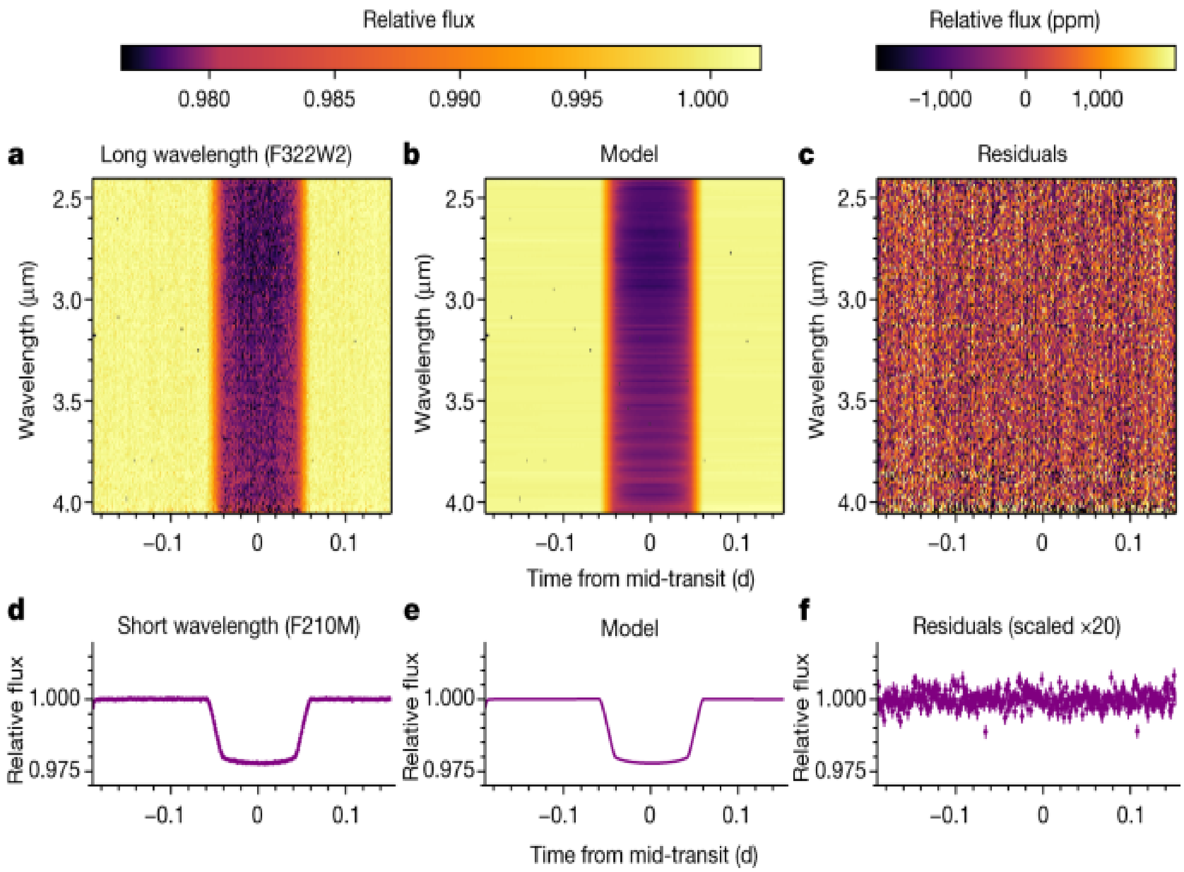
We observed a single transit of WASP-39b with JWST's NIRCam instrument on 22–23 July 2022 (19:28–03:40 ut). The Grism R and F322W2 filter in the long-wavelength (LW) channel dispersed light from 2.420–4.025 μm at a spectral resolution *R* of 1,570–2,594 over 1,023 resolution elements. The short-wavelength (SW) channel allowed the simultaneous measurement of light, that is photometry, spanning 2.0–2.2 μm using the WLP8 weak lens and F210M filter. See Methods for more details.

The team conducted three independent reductions of the NIRCam LW spectroscopic data and four independent fits and analyses of the reduced data. We also performed two independent analyses of the SW photometric

data. For both data reductions (LW and SW), customizing the JWST Science Calibration Pipeline (jwst) to allow for minor adaptations to default steps and values worked best (Methods). The wavelength solution available with the reference files provided by the JWST Calibration Reference Data System at the time of our analysis was inaccurate (particularly for the blue edge of the LW channel), so we redefined our wavelength values using a polynomial wavelength calibration derived from a planetary nebula observed as part of commissioning (programme 1076).

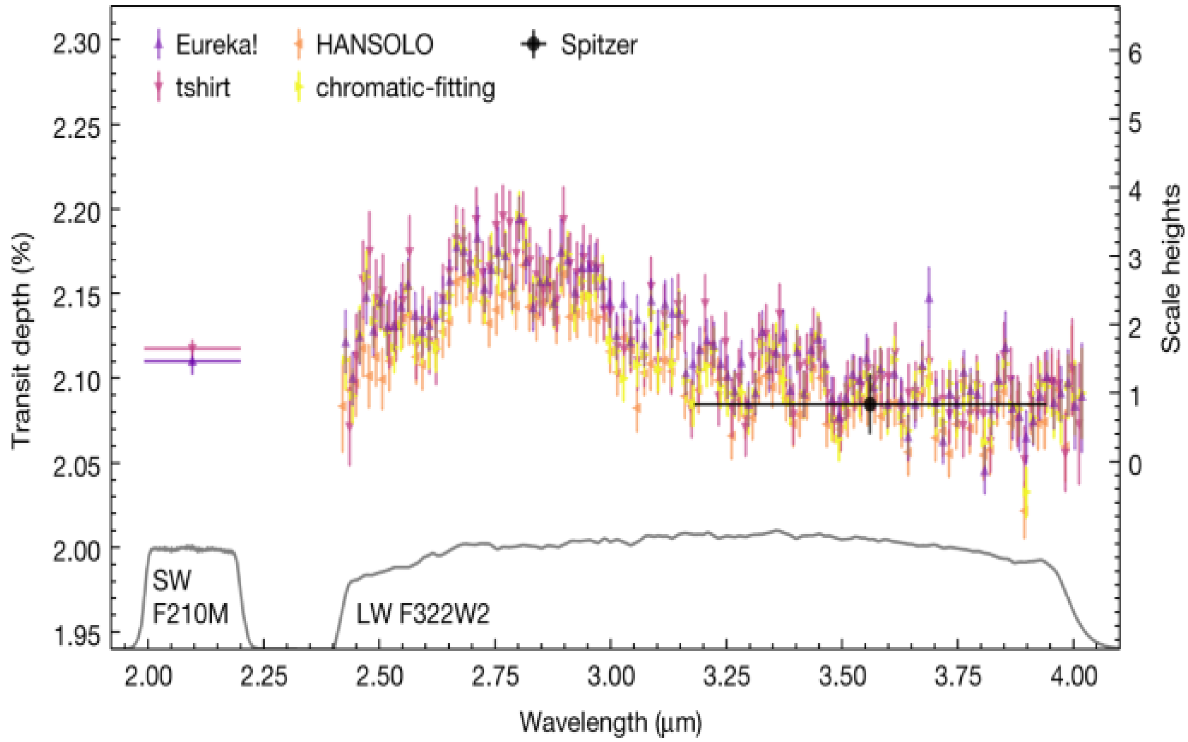
We found no large systematic structures affecting the LW light curves and a minuscule ramp at the start of the SW light curve, see Fig. 1. The only other systematic identified was $1/f$ noise (or pink noise; where f is frequency), which describes the detector's correlated read noise[26]. For NIRCam, this manifests as weak structures in the dispersion direction, as shown in Fig. 1c. We did not correct for $1/f$ noise in the final LW reduction because it did not impact the precision reached by individual spectroscopic light curves (compare tshirt and Eureka! in Fig. 2 for analyses with and without $1/f$ noise corrections). We removed structures due to $1/f$ noise in the SW reduction (Methods). We found that a linear model in time was sufficient to detrend the data, which produced uncertainties 1.18× the photon noise limit (median of 135 ppm for the transit depths) at a binned spectral resolution of 15 nm (about 15 pixels). Similarly, the photometric transit-depth precision was 1.35× the noise limit at 53 ppm. The residuals are Gaussian (Extended Data Fig. 5).

Fig. 1: The relative brightness of the WASP-39 planetary system as a function of time and wavelength, as measured by NIRCam.



a–**f**, Spectroscopic data (**a**–**c**) and the photometric (SW) channel (**d**–**f**) for the extracted flux normalized by the median stellar spectrum (**a**,**d**), the best-fit transit and systematic models (**b**,**e**) and the residuals (**c**,**f**). The flux decrease results from the transit of exoplanet WASP-39b in front of its star. The subtle variation in transit depth around 2.8 μm is due primarily to water vapour in the planet's atmosphere. The vertical striping in the residuals is due to 1/*f* noise.

Fig. 2: The transit spectrum of WASP-39b as measured from JWST's NIRCam instrument.



The coloured points with 1σ uncertainties depict our independent analyses of the spectroscopic LW channel (2.420–4.025 µm) and photometric SW channel (2.0–2.2 µm) with their respective throughputs shown in grey. All analyses agree with the broadband Spitzer point (black circle, 3.2–4.0 µm). The broad feature centred at 2.8 µm spans 2.5 scale heights (∼2,000 km) and is due primarily to water vapour within WASP-39b's atmosphere. We note the consistency between analyses in the fine structure.

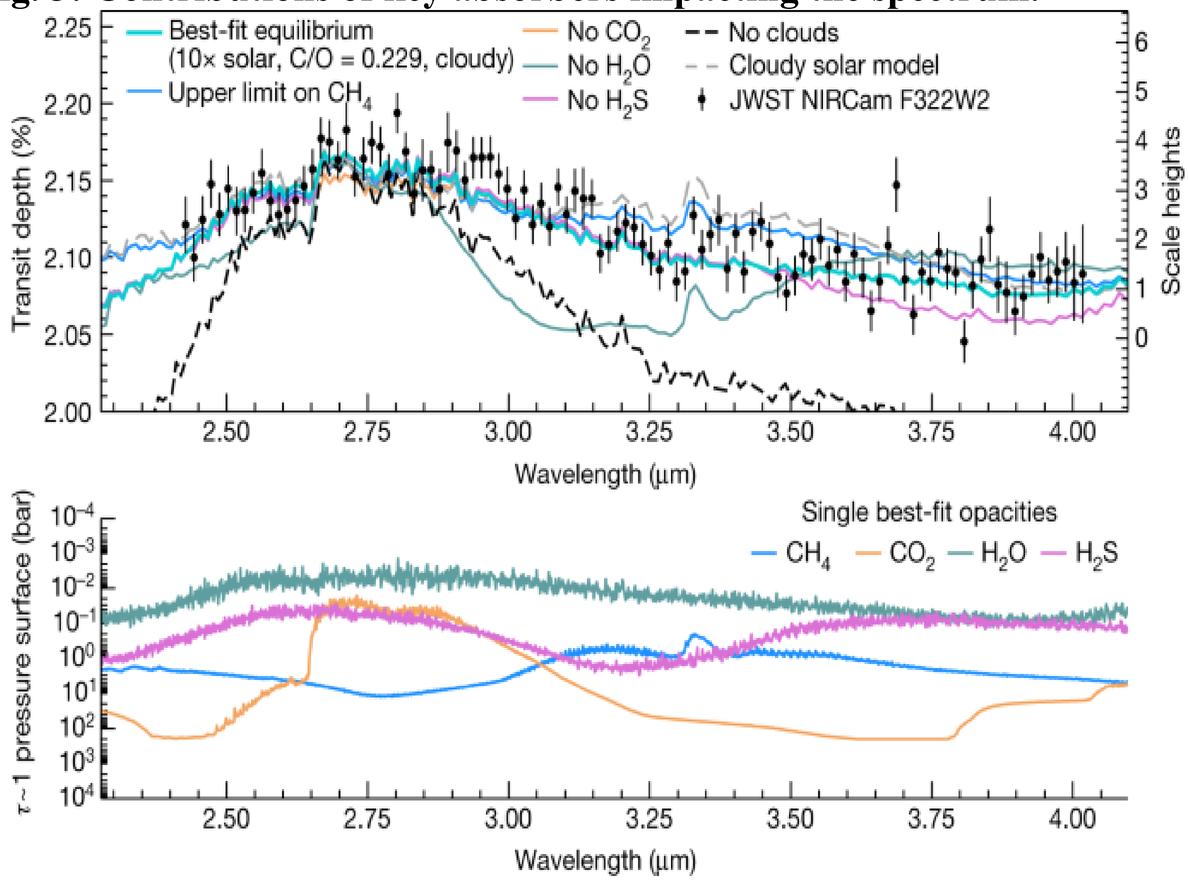
Source data

Figure 2 shows the independently derived transit spectra and photometry. Each reduction is consistent with our selected reduction (Eureka!) to better than 1σ, as is the broadband 3.6-µm Spitzer point[13]. The overall shape of the spectrum is due primarily to absorption of water vapour (feature centred at 2.8 µm). The right-axis scale is in equivalent scale heights, where one scale height is approximately 800 km.

To interpret the presence of other molecules within the planetary atmosphere, we compared the Eureka! transit spectrum with a set of independently computed atmospheric model grids that spanned a range of cloud properties, metallicity values and carbon-to-oxygen (C/O) ratios

(Methods). Figure 3 shows a representative best-fit model highlighting the contributions of major molecular absorbers.

Fig. 3: Contributions of key absorbers impacting the spectrum.



Top: the best-fit PICASO 3.0 equilibrium model (10× solar, C/O = 0.229, moderate grey clouds with cloud optical depth of 2.5×10^{-3}) is shown compared with the Eureka! reduction, along with models with individual molecular species removed to show its contribution to the spectrum. Each model is normalized to the data for illustration by offsetting each model to have the same transit depth at 2.8 μm. Water predominately sets the shape of the spectrum, followed by the influence of clouds. The grey dashed line shows a cloudy solar-metallicity and stellar-C/O atmospheric model, illustrating the lack of a strong CH_4 peak seen in the data. Bottom: the opacities of the dominant molecular species at an optical depth (τ) of 1 in the atmosphere. In the single best-fit model shown in the bottom panel, the CH_4 peak at 3.3 μm is blended out by water absorption. However, manual scaling

of CH_4 gives an upper limit of CH_4 abundance (blue line) for the single best-fit model shown in the top panel.

Source data

Our spectroscopic wavelength range covered by NIRCam/F322W2 includes absorption features due to prominent atmospheric molecules such as water (H_2O), carbon dioxide (CO_2) and methane (CH_4). From our model grid search, we definitively confirm the presence of H_2O at nearly 16σ. Water vapour was previously identified in the atmosphere of WASP-39b using the Hubble Space Telescope (HST) Wide Field Camera 3 (WFC3) observations taken at shorter wavelengths ($[H_2O] = -{1.37}_{-0.13}^{+0.05}$)[13]. We also see weak evidence for CO_2 absorption, previously seen with high confidence using the PRISM mode on the Near InfraRed Spectrograph (NIRSpec) at 4.3 µm (ref. [7]), but the overlap between the CO_2 feature at 2.8 µm and the broad H_2O feature (illustrated in Fig. 3) leads to a more tentative identification here. Each forward model grid prefers significant cloud coverage, which impacts the spectrum at approximately millibar pressures, despite differing cloud parameterizations between grids with varying levels of physical complexity (Methods).

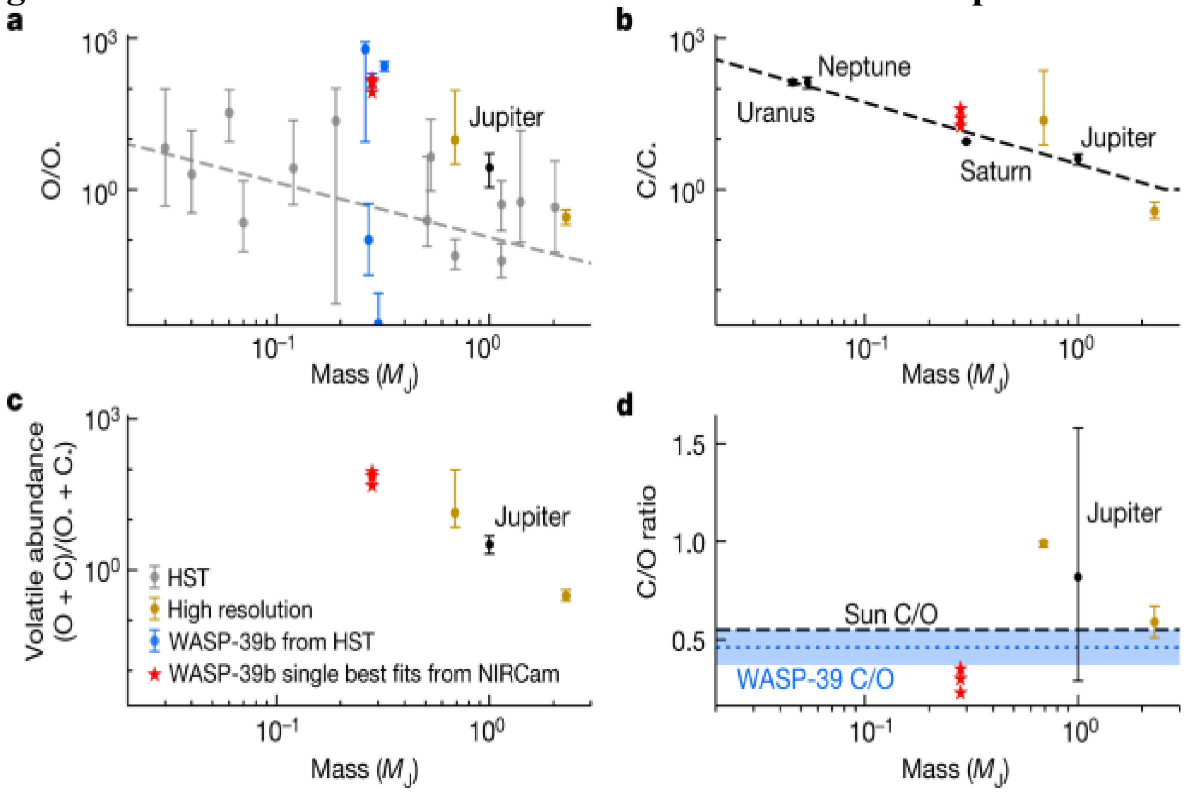
In a hot (about 1,000 K) solar-metallicity atmosphere with a stellar C/O ratio, CH_4 would be visible as a strong peak at 3.3 µm (grey dashed line in Fig. 3, and Extended Data Fig. 7) under thermochemical equilibrium. Such a peak is absent in the reduced spectrum. We quantified this using a residual fitting test (Methods). In a higher-metallicity and/or lower-C/O atmosphere, carbon is increasingly partitioned into CO and CO_2, and the CH_4 peak at 3.3 µm disappears. Therefore, the absence of a strong CH_4 peak at 3.3 µm in our data drives the metallicity to higher values and the C/O ratio to lower values. We scaled the CH_4 volume mixing ratio within our single best-fit Planetary Intensity Code for Atmospheric Spectroscopy Observations (PICASO) version 3.0 model (10× solar metallicity; C/O ratio of 0.229) to determine an upper limit on the abundance of CH_4 at 1 mbar, where it contributes most strongly to the spectrum. Within our single best-fit model scaling, we find an upper limit on CH_4 abundance at 1 mbar of 5.5×10^{-5} (or 55 ppm) volume

mixing ratio, above which the goodness of fit per free parameter, ${\chi }_{\nu }^{2}$, gets increasingly worse (that is, ${\chi }_{\nu }^{2} > 2$). We also tested whether other data reductions favoured best-fit models with stronger CH_4 abundances, but found they did not have any statistical significance.

Driven by this CH_4 upper limit, the single best fit from each grid favours the lowest C/O ratio (0.229, 0.3 and 0.35 for PICASO 3.0, PHOENIX and ATMO, respectively) within that grid. These best-fit point values for C/O from the three grids agree well with the value of ${0.31}_{-0.05}^{+0.08}$ found by ref. [13]. We examined the effect of an even lower C/O grid point by computing the best-fit PICASO 3.0 model with a C/O of 0.115, but found no discernible difference in the transit spectrum. Comparing our inferred C/O ratio for WASP-39b's atmosphere with that of its host star, we see that it is substellar (≤0.35, whereas WASP-39 is 0.46 ± 0.09 (ref. [23])). We also note that the C/O ratio shown here represents the C/O fraction of the planet's upper atmosphere rather than that of the whole atmosphere, as these NIRCam observations probe approximately the 0.1–10 mbar pressure range. WASP-39b's temperature–pressure profile is cool enough for the formation of silicate (that is, Obearing) cloud species at depth, which would deplete oxygen from the upper atmosphere and actually increase the C/O ratio aloft compared with the bulk planetary envelope[27,28].

Figure 4 compares our best-fit metallicity values, shown as separate O and C abundances, and C/O ratios to previous studies using HST data, as well as results for exoplanets observed at high resolution and Solar System gas giants. The JWST/NIRCam data rule out a super-stellar C/O ratio for WASP-39b. In addition, Fig. 4 demonstrates the capability of JWST to measure the C/O ratios of giant planet atmospheres by observing both O- and C-bearing species, which until now has only been achieved through high-resolution exoplanet observations (for example, refs. [29,30]). Similar measurements have been difficult to achieve from HST alone. Even in the Solar System gas giants, such constraints have proved difficult from both remote sensing and in situ missions, as the low temperatures of Jupiter, Saturn, Uranus and Neptune lead to condensation of most O-bearing species (for example, H_2O and CO_2) at high altitudes, prohibiting accurate measurement of the O abundance (for example, refs. [31,32]).

Fig. 4: Trends in elemental abundances and C/O ratio with planet mass.



a–**c**, The abundances of O (**a**), C (**b**) and net volatiles (O + C) (**c**) scaled to stellar values (O_* and C_*). The grey points in **a** show HST constraints based on ≥2σ H_2O detections, with the grey dashed line showing the best-fit trend from ref. [18]. The blue points show all previous estimates of the metallicity of WASP-39b from HST data, offset in mass for clarity[13,15,16,17,18]. The black points and dashed line in **b** show a fit based on CH_4 abundances of Solar System giant planets[46,47,48,49]. Of the Solar System planets, only Jupiter has a constrained O abundance (from Juno observations of H_2O (ref. [32])). The gold points indicate high-resolution observations of H_2O and CO in exoplanets[29,30], and the red stars show the best-fit values for WASP-39b as measured by JWST/NIRCam for each of the three model grids described in this paper. **d**, The black dashed line depicts the solar C/O ratio of 0.55 (ref. [50]) and the blue dotted line with a shaded 1σ uncertainty region indicates the measured C/O ratio of the star WASP-39[23]. Our results for WASP-39b favour a super-stellar volatile abundance and substellar C/O ratio. However, we emphasize that a full retrieval will be necessary to determine accurate means and 1σ error bars for the NIRCam results.

Source data

The apparent substellar C/O ratio inferred from chemical equilibrium models may trace photochemical processes in the planet's upper atmosphere. For example, photochemical destruction of CH_4 in the upper atmosphere could explain the absence of a CH_4 peak at 3.3 μm (for example, refs. [6,33]). The most likely immediate products of CH_4 photolysis, such as hydrogen cyanide (HCN) or acetylene (C_2H_2), would be produced in abundances too small (less than or approximately equal to a few parts per million[6,33]) to be robustly detected with a single NIRCam transit, even from complete CH_4 conversion. Alternatively, much of the C available from CH_4 photolysis could have been oxidized by photodissociated H_2O to form carbon monoxide (CO) and CO_2 (refs. [6,33,34,35]), although the absolute abundances of these two C reservoirs would not have been meaningfully altered as their abundances under chemical equilibrium are already higher than that of CH_4. Other proposed disequilibrium chemistry processes could reduce the CH_4 abundance at the terminator without also decreasing the C/O ratio[5,36,37,38,39]. We defer the exploration of complex disequilibrium models to atmospheric retrieval analyses using the full set of data provided by the Transiting Exoplanet Community ERS programme. That dataset will also constrain the presence of additional O- and C-bearing species to provide a more robust constraint on the C/O ratio than we can obtain here. However, the C/O ratio estimate we report from NIRCam is broadly consistent with the C/O ratio found from the other individual ERS WASP-39b datasets, which range from best fits that are subsolar (Near InfraRed Imager and Slitless Spectrograph (NIRISS)/Single Object Slitless Spectroscopy (SOSS)[40]; NIRSpec/PRISM 3.0–5.0 μm (ref. [7]); NIRSpec/G395[41]) to a slightly super-solar upper limit (NIRSpec/PRISM 0.5–5.5 μm (ref. [42])).

If disequilibrium chemistry is not prevalent in the planet's upper atmosphere, the inferred high metallicity and low C/O ratio can be tied back to the formation of WASP-39b. The most prominent scenario is that WASP-39b formed by core accretion exterior to the water-ice line and accreted low-C/O solid material in situ and/or while migrating inwards within the protoplanetary disk[4,43,44]. Taken as such, JWST observations could offer

important clues regarding the degree to which hot-Jupiter atmospheres undergo solid accretion during their early evolution.

Here we have demonstrated the excellent performance of NIRCam for exoplanet transmission spectroscopy. With the first JWST exoplanet spectra now comparable to the first near-infrared Jupiter spectra[45], the future promises many exciting discoveries and major advancements in the formation, evolution and atmospheric chemistry of hot Jupiters.

Methods

As part of this article's Reproducible Research Compendium, located on Zenodo at https://doi.org/10.5281/zenodo.7101283, we provide saved outputs from various pipeline stages and the data used to generate relevant figures, as well as a Jupyter Notebook with step-by-step data reduction instructions replicating our chosen analysis (https://doi.org/10.5281/zenodo.7510106).

Photometric monitoring of host star

To confirm that WASP-39 is a relatively inactive star, and that the JWST observations were not adversely affected by stellar activity, we carried out photometric monitoring with the ground-based NGTS[11]. Monitoring began at the end of April 2022 and continued until late August, spanning the JWST ERS transit observations of WASP-39b in July. We used one camera on most photometric nights to take a series of 10-s images lasting on average for 2 h. The resulting monitoring light curve is plotted in Extended Data Fig. 1 (top), showing one binned point for each night. Also included is the TESS sector 51 Pre-search Data Conditioned Simple Aperture Photometry (PDCSAP) light curve of WASP-39[12], which is binned to 2 h to be comparable to NGTS. Both light curves have been detrended against sky brightness. They show evidence for stellar activity, but only with a low amplitude of 0.06% in NGTS.

Also plotted in Extended Data Fig. 1 (bottom) are individual transit observations of WASP-39b with NGTS and TESS (the times of which are indicated on the monitoring light curve). For four of the NGTS transits, we

employed multiple cameras. This significantly improves the photometric precision[51], which is otherwise limited by atmospheric scintillation[52]. The transit models were generated from the system parameters listed in Extended Data Table 1. We fit only the transit times and the mutual depth of the TESS transits, which is slightly shallower than expected.

The transit observations in Extended Data Fig. 1 show no evidence for starspot-crossing events, which would be visible as bumps in the transit light curve. The absence of such events across multiple high-precision transits provides additional evidence that WASP-39 is a quiet star and that the JWST ERS transit observations are unlikely to be adversely affected by stellar variability.

JWST NIRCam observation

JWST observed the 2.8-h transit of WASP-39b over a span of 8.2 h, providing a baseline before and after transit to measure transit depths accurately. A dichroic beam splitter allows NIRCam to simultaneously observe a target in both SW and LW channels[53,54]. The LW channel used the Grism R + F322W2 filter to observe a wavelength range of 2.420–4.025 μm with a spectroscopic resolving power of $R \approx 1,600$ at 4 μm (Extended Data Fig. 1, top). The SW imaging channel used the WLP8 weak lens and F210M filter (2.0–2.2 μm) to produce the hexagonal pattern shown in Extended Data Fig. 2 (bottom). Spreading the light prevents saturation, reduces variability owing to image motion over an imperfect flat field and allows monitoring of mirror-segment alignment. Both SW and LW channels used the SUBGRISM256 subarray mode with four output amplifiers and the SHALLOW4 readout pattern to minimize data volume. With 12 groups per integration (82.17 s total), we acquired 366 integrations for this transit observation.

Data reduction and calibration

We conducted independent data analyses using multiple pipelines and fitting tools to ensure that we obtained the same transmission spectrum using different reduction pipelines. We also varied the fitting methods within a given data reduction pipeline.

Many of the reductions presented below used intermediate data products from or made minor edits to the JWST Science Calibration Pipeline (jwst; https://jwst-pipeline.readthedocs.io/), which we briefly summarize here. jwst is a Python software suite for processing data from all JWST instruments and observing modes, and is organized into three stages. Stage 1 takes in uncal.fits files and performs detector-level corrections and ramp fitting for individual exposures (that is, ramps-to-slopes conversion; these ramps are the flux increases during an exposure, not to be confused with baseline ramps over the course of the entire transit). Stage 2 takes in slope images (ramps) from Stage 1 and performs an assignment of the world coordinate system, flat fielding and assignment of a wavelength solution. Stage 3 takes in calibrated two-dimensional images from Stage 2 and extracts a time series of one-dimensional spectra. The default pipeline settings include a flux calibration step at Stage 2. In all data reductions presented below, we skipped that step, as it introduced scatter in the extracted spectral time series. This is justified because the transit depths we compute are relative, rather than absolute, flux measurements.

Below we describe the independent data reductions applied to the SW photometry and LW spectroscopy, respectively. In each case, we note where data reductions deviated from the standard jwst pipeline.

SW photometry

We performed two independent SW data reductions using the open-source Eureka! and tshirt pipelines.

Eureka! SW reduction

Eureka! is an open-source pipeline designed to perform spectral extraction and fitting for JWST exoplanet time-series observations[55]. The Eureka! SW data reduction used the default jwst settings for stages 1 and 2, with the exception of increasing the rejection threshold during jump detection to 10σ, which improved the quality of the resulting light curve.

In Stage 3, we first masked all pixels for which the 'DO_NOT_USE' data quality flag was raised by the jwst pipeline. We then performed an outlier

rejection along the time axis for each individual pixel in a segment using a 7σ threshold, repeating this process twice. Next, we corrected for the $1/f$ noise in each of the four amplifier regions by subtracting the median flux in each row calculated without pixels containing the star. We interpolated over flagged pixels using a cubic function. Finally, we determined the image centre and performed aperture photometry on the target. We explored different target apertures and background annuli, and chose the combination that minimized the root-mean-square variations, leading to a target aperture radius of 65 pixels and a background annulus from 70 pixels to 90 pixels relative to the centre.

tshirt SW reduction

tshirt is an open-source pipeline (https://tshirt.readthedocs.io/en/latest/) that has tools to modify the jwst pipeline and performs photometric and optimal spectral extraction of light curves.

In the stage 1 SW analysis, tshirt applied a row-by-row, odd/even-by-amplifier (ROEBA) subtraction algorithm that used background pixels to reduce the $1/f$ noise. In this procedure, background pixels are used to correct each group in a similar fashion to reference pixel correction (https://jwst-pipeline.readthedocs.io/en/latest/jwst/refpix/index.html). The ROEBA correction happens after the bias subtraction step. First, the median of all even columns' background rates is subtracted from all even columns and the median of all odd columns' background rates is subtracted from all odd columns to remove most pre-amp reset offsets and odd/even pixel effects. Next, the median of each column's background rate is subtracted from each row to remove the $1/f$ noise for timescales longer than a row read time (5.24 ms). The correction was applied to each group so that $1/f$ noise would not be detected as spurious jumps or cosmic rays by the pipeline. We used all pixels more than 201 pixels from the source to estimate the background and $1/f$ noise, then subtracted the median of each row from all pixels in that row. Stage 2 of jwst was skipped, as it only changes the rates from analogue-to-digital units (ADU) per second to physical units and conducts flat fielding. This does not affect the relative measurements of the light curve (due to the

high pointing precision) and allows for comparison with detector-level effects.

For the photometric extraction, we used a source radius of 79 pixels and a background annulus of 79 pixels to 100 pixels. We performed a two-dimensional Gaussian fit to determine the centre of the aperture.

LW spectroscopy

We performed three independent LW data reductions, using the Eureka!, Atmospheric Transmission Spectroscopy Analysis Code (HANSOLO) and tshirt pipelines.

The reference files in the Calibration Reference Data System at the time of our analysis included a linear solution for wavelength as a function of x coordinate (the dispersion direction), but this is not strictly accurate at the blue end. For all methods, we use commissioning programme 1076 to derive a third-degree polynomial wavelength solution that uses the Pfund and Bracket hydrogen series in the planetary nebula IRAS 05248−7007. The residuals in this solution are $\lesssim$0.1 nm and the stellar absorption lines in WASP-39 agree with the solution to within 1 nm. The difference between the corrected wavelengths and the original wavelength solution is almost zero at the red end of the spectrum, but increases to about 50 nm at the blue end.

Eureka! LW reduction

We investigated several variations of the Eureka! LW data reduction to minimize the median absolute deviation (MAD) of the final extracted light curves, with different settings for cosmic-ray jump detection, identifying the spectral trace, the aperture size for spectral extraction, the region for background subtraction and limits for outlier rejection. Here we present details of the data reduction that produced the spectrum shown in the main body of the paper.

Stages 1 and 2 were identical to the jwst pipeline, with the exception of increasing the rejection threshold during jump detection to 6σ. In Stage 3,

we first trimmed the data to a subarray extending from pixels 4–64 in the cross-dispersion direction and 4–1,704 in the spectral direction. We then masked any pixels with not a number (NaN) values for the flux or error. We fit the spectral trace with a Gaussian profile and corrected for the curvature of the trace to the nearest integer pixel. We excluded a 14-pixels-wide region on either side of the spectral trace from the background calculation and performed a column-by-column linear fit to subtract the background. We used a double-iteration 7σ threshold for outlier rejection of the sky background along the time axis during background subtraction. In addition, we used a 7σ threshold for outlier rejection during the polynomial fit to the background. To obtain the spectrum, we constructed a normalized spatial profile using the median of all data frames, then used optimal extraction[56] on an aperture with a half-width of 9 pixels. For the optimal extraction, we rejected outliers above a 10σ threshold. Extended Data Fig. 3 shows the curvature-corrected, background-subtracted median frame with indicated background and aperture regions.

HANSOLO LW reduction

The HANSOLO pipeline was originally developed to analyse ground-based transmission spectra observed with 8-m-class telescopes[57,58] and was adapted to enable its use on NIRCam data. HANSOLO begins with the calibrated rateints.fits outputs of jwst Stage 1.

We used the LACOSMIC algorithm[59] to remove cosmic-ray effects from the two-dimensional images and identified the spectral trace using a Moffat function fit to each column. To remove the sky, we fitted and subtracted a linear trend from each column, excluding from the fit a region of 20 pixels on either side of the trace centre. We then extracted the spectrum by summing over an aperture with a half-width of 3 pixels. The spectra from different images were aligned with each other using cross-correlation. To correct outlier pixels, each spectrum was normalized to account for the effect of the transit on the time series. Outliers >3σ away from the mean were removed from the time series of each wavelength point in the normalized spectra and replaced with the median value over time. We then rescaled the spectra to their original amplitudes.

tshirt LW reduction

As with the SW reduction, a few modifications were made to the Stage 1 jwst ramps-to-slopes pipeline. ROEBA subtraction reduced 1/*f* noise (described above for photometry); however, only pixels 1,847 to 2,044, which are on the rightmost amplifier, are available as low-illumination background.

For Stage 3, tshirt performed optimal spectral extraction weighted by the covariance between pixels[26]. We used a spectral aperture centred at pixel 34 in the spatial direction with a half-width of 5 pixels. We selected the background region to extend between pixels 5–24 and 44–65 in the spatial direction. The background was fit with a column-by-column linear trend with 3σ clipping. For the spectral extraction, we fit the spatial profile with a cubic spline with 20 knots and an outlier rejection threshold of 30σ. If a pixel was deemed an outlier either by the 'DO_NOT_USE' data quality flag or by the spatial profile outlier detection, the rest of the spatial profile was weighted by the reference profile to ensure that the flux was conserved. For the covariance weighting, a correlation of 8% was assumed between pixels as measured by background pixels' noise properties.

Data analysis and fitting

We used both Eureka! and tshirt to fit the SW light curves. In both cases, the light curves were fit with models that included both the transit and the systematic noise. However, to investigate the effect of different systematic models on the resulting spectra, each fit used a slightly different noise model. Extended Data Table 1 summarizes the systematics models that were used in each SW fit.

For the LW fits, we summed the data into 15 nm bins (about 15 pixels). We experimented with bins as small as 10 nm, but found that reducing the bin size below 15 nm led to poor constraints on the limb darkening and added additional scatter to the resulting spectrum. Extended Data Fig. 4 shows that the noise is primarily Gaussian out to long timescales of order the length of ingress/egress. In addition, we created a white-light curve by summing the extracted spectra over the entire 2.420–4.025 μm wavelength region. We

experimented with different wavelength cut-offs but chose to extract spectra in this wavelength region because the low instrument throughput affected the quality of the extracted light curves beyond this region. Extended Data Fig. 5 shows all reduced transmission spectra with one bin added on the blue end and two added on the red end, as well as the relative throughput at the wavelengths of these bins. This figure shows the large error bars derived from data near the edges of the NIRCam/F322W2 bandpass. Therefore, we recommend that future works limit extracted spectra to the wavelength region between 2.420 μm and 4.025 μm.

We fit the LW light curves using four independent pipelines: chromatic-fitting, Eureka!, HANSOLO and tshirt. chromatic-fitting is an open-source (https://github.com/catrionamurray/chromatic_fitting/) Python tool to perform light-curve fitting, built on the data visualizer chromatic (Z. K. Berta-Thompson, manuscript in preparation; https://github.com/zkbt/chromatic/). For this work, chromatic-fitting light-curve fitting was applied to a Eureka! data reduction. As with the SW fits, we fit the LW light curves with models that include different noise parameterizations. Extended Data Table 2 summarizes the systematics models that were used in each LW fit.

For all fits, the parameters were estimated with a Markov chain Monte Carlo fit, using either the emcee Python package[60] (for fits performed with Eureka!), the pymc3 Python package[61] (implemented through the Exoplanet code[62,63], for fits performed with chromatic-fitting or tshirt) or the CONAN Python package[57,58] (for fits performed with HANSOLO). The number of free parameters and the resulting differential MADs of the light curves from each fit are also listed in Extended Data Tables 1 and 2. The best-fit parameters from the white-light-curve fits are given in Extended Data Table 3.

In the process of performing the fits to the LW data, we regularly found that the best-fit transmission spectra were shifted vertically for different limb-darkening parameterizations and, for some reductions, exhibited changes in the apparent size of the water feature. In particular, we found that light-curve fits with all limb-darkening coefficients fixed to outputs from ExoTiC-LD[64,65,66] could result in a biased planet spectrum and might present a

higher level of time-correlated noise in the residuals. We attribute this to a combination of JWST's high-precision light curves and deficiencies in the stellar limb-darkening models to accurately represent WASP-39[67,68]. Therefore, the results presented here use the quadratic limb-darkening law, in its classical form or reparameterized by ref. [69], with one or both coefficients as free parameters. We confirmed that these parameterizations produce transmission spectra that are consistent both with each other and with the spectra resulting from using more complex limb-darkening parameterizations, such as a four-parameter law with either fixed or free parameters[70]. We therefore recommend that future transmission spectrum analyses with NIRCam use similar methods. Limb-darkening conclusions from the full Transiting Exoplanet Community ERS programme will be discussed further by N. Espinoza et al. (manuscript in preparation).

The final fitted light curves are shown in Extended Data Fig. 6 and the final transmission spectra are shown in Fig. 2. Both the SW and LW datasets are also available in our Reproducible Research Compendium on Zenodo at https://doi.org/10.5281/zenodo.7101283. The median difference between each transmission spectrum and the Eureka! spectrum is 0.87σ (using the maximum error at each point), which demonstrates a remarkable level of agreement. In addition, the residuals showed no evidence for time-correlated noise, as shown in Extended Data Fig. 5.

For ease of interpretation, we compared our atmospheric models with only one transmission spectrum. We selected the Eureka! spectrum, as it was on average nearest the median spectrum (the median transit depth at each bin).

Atmospheric forward modelling

To interpret the LW data from NIRCAM/F322W2, we performed χ^2 fits to the transmission spectra using three grids of radiative–convective equilibrium models: ATMO[71,72,73], PHOENIX[74,75,76] and PICASO 3.0[77,78]. All models used a common set of planetary parameters, but had differing opacity sources, cloud treatments and grid points, described in detail below. Each model was binned to the resolution of the data to perform the χ^2 fitting. We performed these three independent model grid fits to fully vet our

inferences about the atmospheric metallicity and the presence of specific molecular features within the data.

The PICASO 3.0, Vulcan and Virga model grid

Our primary atmospheric model grid is built from the open-source radiative–convective equilibrium code PICASO[77], version 3.0[78], which was developed from the Fortran-based Extrasolar Giant Planet (EGP) model[79,80,81]. We used PICASO 3.0 to generate one-dimensional temperature–pressure profiles in thermochemical equilibrium. The base PICASO 3.0 forward model grid computes atmospheric mixing ratios using variations of planetary intrinsic temperatures (T_{int}) of 100 K, 200 K and 300 K; C/O ratios of 0.229, 0.458, 0.687 and 0.916; and atmospheric solar metallicity values of 0.1×, 0.316×, 1.0×, 3.162×, 10.0×, 31.623×, 50.119× and 100× solar. The PICASO grid assumes full day–night energy redistribution. To compute model transmission spectra from the atmospheric profiles, we used opacities described by ref. [81] (see in particular Extended Data Table 2), which sources H_2O from refs. [82,83], CH_4 from refs. [84,85,86], CO_2 from ref. [87] and hydrogen sulfide (H_2S) from refs. [84,88,89].

We then used the one-dimensional CHON-based chemical kinetics code VULCAN[33] and the cloud modelling code Virga[90], which is the Python implementation of the Eddysed cloud code[91], to post-process disequilibrium chemistry from mixing and photochemical products as well as the effect of clouds. These additional post-processed grids also include vertically constant eddy diffusivities (K_{zz}) of 10^5–10^{11} cm^2 s^{-1} in steps of 2 dex, and both clear and cloudy models. For the Vulcan disequilibrium runs, we computed model grid points for only a select subset of metallicity values (1×, 10×, 50× and 100× solar) and C/O ratios (0.229, 0.458 and 0.687). We found that neither the cloudy nor the clear disequilibrium grids from VULCAN offered an improvement in the \({\chi }_{\nu }^{2}\) value. Given the sparseness of these pre-computed disequilibrium grid models, we left rigorous quantification of self-consistent disequilibrium chemistry in the atmosphere of WASP-39b to future work.

Within PICASO, clouds are implemented both as grey absorbers and as Mie scatterers using temperature-relevant cloud condensate species from Virga. For the grey clouds, the grid specified a cloud optical depth (τ_{cloud}) between 1 bar and 0.1 bar ranging from $\tau_{cloud} = 3.2 \times 10^{-6}$ to 1 in steps of 0.1 dex across all wavelengths. For clouds of specific condensates, we used Virga to compute log-normal particle size distributions using sedimentation efficiency (f_{sed}) values of 0.6 to 10 for MnS, Na_2S and $MgSiO_3$ along the range of K_{zz}. Smaller sedimentation efficiencies, f_{sed}, with larger eddy diffusivities, K_{zz}, generated more extended cloud decks and stronger cloud opacity.

The PHOENIX model grid

We also used a grid of atmosphere models from the PHOENIX radiative–convective equilibrium code to fit the data[74,75,76]. Similar to the PICASO 3.0 grid, parameters including the day–night energy redistribution factors, interior temperature (200 K and 400 K), bulk atmospheric metallicity (0.1×, 1×, 10× and 100× solar) and C/O ratio (136 grid points from 0.3 to 1) were varied. Aerosol properties were parameterized through a haze factor (0 and 10× multi-gas Rayleigh scattering) and a grey-cloud-deck pressure level (0.3 mbar, 3 mbar and 10 mbar). Models with molecular abundances quenched at 1 bar to simulate vertical mixing were also calculated. The grid also included rainout to account for species sequestered as condensates in the deep atmosphere. Opacities are described by refs. [76,92] and taken from ref. [88].

The ATMO model grid

Similar to the model grids described above, we compared the data to a grid of models from the ATMO radiative–convective–thermochemical equilibrium code[71,72,73,93]. The ATMO grid used similar atmospheric and aerosol parameterizations to those used in the PHOENIX grid and also included rainout that accounts for species condensed in the deep atmosphere. Also included are day–night energy redistribution factors (0.25, 0.5, 0.75 and 1; with 1 as full redistribution), atmospheric metallicity (0.1×, 1×, 10×

and 100× solar), interior temperature (100 K, 200 K, 300 K and 400 K), C/O ratio (0.35, 0.55, 0.7, 0.75, 1.0 and 1.5), cloud scattering factor (0, 0.5×, 1×, 5×, 10×, 30× and 50× H_2 Rayleigh scattering at 350 nm between 1 mbar and 50 mbar pressure levels) and a haze scattering factor (1× and 10× multi-gas Rayleigh scattering). Opacities for H_2O, CO_2 and CH_4 are taken from refs. [83,84,85,86] and for H_2S from ref. [88].

Grid fits to JWST/NIRCam data

We applied each of our three grids—ATMO, PHOENIX and PICASO 3.0—to fitting the NIRCam F322W2 spectrum (2.4–4.0 µm). In doing so, we found that the models strongly favoured a solar- or super-solar-metallicity atmosphere (1–100× solar), a substellar C/O ratio (≤0.35) and substantial contribution from clouds, which are parameterized differently by each model grid (see each grid description above). We show the best fits from each model grid in Extended Data Fig. 7. This interpretation is in agreement with the results using JWST's NIRSpec/PRISM instrument from 3.0–5.0 µm (ref. [7]), improving on the wider spread from previous HST-only[13,14,15,17,18,94] or HST and ground-based optical interpretations[16].

For the NIRCam-only fit, the PICASO grey-cloud scheme produced a slightly better best fit (\({\chi }_{\nu }^{2}\) = 1.16) than the PICASO + Virga more realistic clouds (\({\chi }_{\nu }^{2}\) = 1.23), both of which were preferred to the clear-model best fit (100× solar) with \({\chi }_{\nu }^{2}\) = 1.53. The Virga best-fit grid resulted in an atmosphere of 1× solar metallicity, C/O = 0.229, f_{sed} = 0.6 and K_{zz} = 10^9 cm s^{-2}. This Virga best-fit model consists of clouds of MnS and $MgSiO_3$ with deep (≥100 bar) cloud bases and diminishing optical depth up to approximately millibar pressures.

The best-fit equilibrium model from the PHOENIX grid had 100× solar metallicity, a C/O ratio of 0.3 and a cloud deck at 3 mbar. Cloudy models were generally preferred over clear models, but not with statistical significance (\({\chi }_{\nu }^{2}\) of 1.25 compared with 1.22). The PHOENIX grid finds best fits with very high metallicity (100× solar), so this low confidence regarding clouds reflects the cloud-metallicity degeneracy

inherent in data restricted to narrow wavelengths (for example, ref. [95]), as well as potentially the sparseness of the model grid.

For the ATMO grid, the best-fit equilibrium model to the NIRCam spectrum was 1× solar metallicity, a C/O ratio of 0.35, a cloud factor of 5 and a haze factor of 1. As with the other two grids, strongly cloudy models (cloud factor of ≥5) were preferred to clear models (\({\chi }_{\nu }^{2}\) of 1.1 versus 1.2).

HST + NIRCam

In Extended Data Fig. 8, we show the comparison between the spectra of HST/WFC3 (G141 and G102, covering 0.8–1.65 μm) and JWST/NIRCam (F210M + F322W2, 2.0–4.0 μm). We chose to show only WFC3 observations from HST, as these are of higher precision than observations from the Space Telescope Imaging Spectrograph or ground-based data[13]. In addition, as HST/WFC3 has the most archival exoplanet data of any instrument on HST, the future JWST exoplanet programme will primarily rely on this HST instrument for inter-telescope comparisons or extending the wavelength coverage of JWST data. For example, the addition of optical and shorter wavelength near-infrared data can help break metallicity degeneracies by better constraining the presence and extent of clouds[13] (for example, ref. [95]). High-altitude clouds or hazes can be inferred from their particle sizes, where small particles scatter shorter wavelengths more efficiently (for example, refs. [96,97]), thus enabling the disentanglement of a very cloudy, low-metallicity atmosphere from a less cloudy, high-metallicity atmosphere[17].

Molecular detections

Once we found the 'single best fit' for the PICASO grid to the NIRCam spectrum (10× solar, C/O = 0.229, grey-cloud optical depth = 2.6×10^{-3} from 1 bar to 0.1 bar), we used this as a base model to explore the significance of specific molecular detections. First, we tested whether we could improve the best fit in the presence or absence of H_2O, CO_2, CH_4 or H_2S. We re-ran the

best-fit base model by zeroing out each of these species in turn, shown in Fig. 3, and then repeating our χ^2 analysis.

We found that although the presence of H_2O, H_2S and CH_4 resulted in a better \({\chi }_{\nu }^{2}\) value, only H_2O and H_2S did so in a statistically meaningful way. As H_2S does not contain strong molecular features within the NIRCam wavelength range, the Gaussian residual fitting we perform for the detection significance of other molecules is not applicable, and we left its further quantification to more rigorous atmospheric retrieval analyses. Increasing the CH_4 abundance beyond that of the best-fit model also improved the \({\chi }_{\nu }^{2}\), although again not to high statistical significance.

With the best fit in hand, we investigated the presence of individual molecular species. For molecular detection significances, we performed the same Gaussian residual fitting, shown in Extended Data Fig. 9, as for the detection of CO_2 in the NIRSpec/PRISM 3.0–5.0 μm analysis[7]. We find a Bayes factor, ln(*B*), of 123.2 between the Gaussian residual and constant models for H_2O over the whole NIRCam wavelength range, corresponding to 15.9σ, a strong detection. For CO_2, we find ln(*B*) of 0.82 between the Gaussian residual and constant models between 2.4 μm and 2.9 μm, or 1.9σ, which is a weak or non-detection[98]. CO_2 is strongly detected at 4.3 μm in the NIRSpec data for WASP-39b[7,41,42], but the strong overlapping H_2O band at 2.8 μm prevents NIRCam from making a significant CO_2 detection. Given our upper limit on CH_4 abundance, we also performed the same Gaussian residual fitting for CH_4 and find a weak or non-detection at approximately 2σ.

Both WASP-39b NIRSpec datasets[7,41,42] observed evidence for a molecular feature near 4.0 μm, which is currently best explained by sulfur dioxide. The reddest data points (>4.025 μm) from NIRCam also show an increase that is consistent with this feature seen in the NIRSpec data. However, as shown in Extended Data Fig. 5, these NIRCam data points have very large error bars because the detector throughput drops off dramatically past 4.0 μm. Future investigations to thoroughly explore the physicochemical likelihood of

sulfur dioxide in the atmosphere of WASP-39b must rely on wavelengths that can fully capture the complete absorption feature, which is beyond the reach of high-fidelity NIRCam/F322W2 measurements.

Data availability

The data used in this paper are associated with JWST programme ERS 1366 (observation 2) and are available from the Mikulski Archive for Space Telescopes (https://mast.stsci.edu). We used calibration data from programme 1076. All the data and models presented in this publication can be found at https://doi.org/10.5281/zenodo.7101283. Source data are provided with this paper.

Code availability

The codes used in this publication to extract, reduce and analyse the data are as follows: batman[99], emcee[60], Eureka![55], jwst[100], chromatic, chromatic-fitting, PyMC3[61], Exoplanet[62,63], gCMCRT[101], CONAN[57,58], ExoTiC-LD[64,65,66], LACOSMIC[59], PICASO[77,78], Virga[90] and VULCAN[33]. The Jupyter notebook to reproduce the transmission spectrum as shown here using the Eureka! code is open-source and can be found at https://doi.org/10.5281/zenodo.7510106.

References

1. Stevenson, K. B. et al. Transmission spectroscopy of the hot Jupiter WASP-12b from 0.7 to 5 µm. *Astron. J.* **147**, 161 (2014).

2. Sing, D. K. et al. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature* **529**, 59–62 (2016).

3. Öberg, K. I., Murray-Clay, R. & Bergin, E. A. The effects of snowlines on C/O in planetary atmospheres. *Astrophys. J. Lett.* **743**, L16 (2011).

4. Espinoza, N., Fortney, J. J., Miguel, Y., Thorngren, D. & Murray-Clay, R. Metal enrichment leads to low atmospheric C/O ratios in transiting giant exoplanets. *Astrophys. J. Lett.* **838**, L9 (2017).

5. Cooper, C. S. & Showman, A. P. Dynamics and disequilibrium carbon chemistry in hot Jupiter atmospheres, with application to HD 209458b. *Astrophys. J.* **649**, 1048–1063 (2006).

6. Moses, J. I. et al. Disequilibrium carbon, oxygen, and nitrogen chemistry in the atmospheres of HD 189733b and HD 209458b. *Astrophys. J.* **737**, 15 (2011).

7. JWST Transiting Exoplanet Community Early Release Science Team. Identification of carbon dioxide in an exoplanet atmosphere. *Nature* https://doi.org/10.1038/s41586-022-05269-w (2022).

8. Stevenson, K. B. et al. Transiting exoplanet studies and community targets for JWST's Early Release Science program. *Publ. Astron. Soc. Pac.* **128**, 094401 (2016).

9. Bean, J. L. et al. The Transiting Exoplanet Community Early Release Science program for JWST. *Publ. Astron. Soc. Pac.* **130**, 114402 (2018).

10. Faedi, F. et al. WASP-39b: a highly inflated Saturn-mass planet orbiting a late G-type star. *Astron. Astrophys.* **531**, A40 (2011).

11. Wheatley, P. J. et al. The Next Generation Transit Survey (NGTS). *Mon. Not. R. Astron. Soc.* **475**, 4476–4493 (2018).

12. Ricker, G. R. et al. Transiting Exoplanet Survey Satellite (TESS). *J. Astron. Telesc. Instrum. Syst.* **1**, 014003 (2015).

13. Wakeford, H. R. et al. The complete transmission spectrum of WASP-39b with a precise water constraint. *Astron. J.* **155**, 29 (2018).

14. Fischer, P. D. et al. HST Hot-Jupiter Transmission Spectral Survey: clear skies for cool Saturn WASP-39b. *Astrophys. J.* **827**, 19 (2016).

15. Tsiaras, A. et al. A population study of gaseous exoplanets. *Astron. J.* **155**, 156 (2018).

16. Kirk, J. et al. LRG-BEASTS: transmission spectroscopy and retrieval analysis of the highly inflated Saturn-mass planet WASP-39b. *Astron. J.* **158**, 144 (2019).

17. Pinhas, A., Madhusudhan, N., Gandhi, S. & MacDonald, R. H_2O abundances and cloud properties in ten hot giant exoplanets. *Mon. Not. R. Astron. Soc.* **482**, 1485–1498 (2019).

18. Welbanks, L. et al. Mass–metallicity trends in transiting exoplanets from atmospheric abundances of H_2O, Na, and K. *Astrophys. J. Lett.* **887**, L20 (2019).

19. Changeat, Q. et al. KELT-11 b: abundances of water and constraints on carbon-bearing molecules from the Hubble transmission spectrum. *Astron. J.* **160**, 260 (2020).

20. Mugnai, L. V. et al. ARES. V. No evidence for molecular absorption in the HST WFC3 spectrum of GJ 1132 b. *Astron. J.* **161**, 284 (2021).

21. Swain, M. R. et al. Detection of an atmosphere on a rocky exoplanet. *Astron. J.* **161**, 213 (2021).

22. Libby-Roberts, J. E. et al. The featureless HST/WFC3 transmission spectrum of the rocky exoplanet GJ 1132b: no evidence for a cloud-free primordial atmosphere and constraints on starspot contamination. *Astron. J.* **164**, 59 (2022).

23. Lodders, K. & Fegley, B. *The Planetary Scientist's Companion* (Oxford Univ. Press, 1998).

24. Kreidberg, L. et al. A precise water abundance measurement for the hot Jupiter WASP-43b. *Astrophys. J. Lett.* **793**, L27 (2014).

25. Fletcher, L. N., Orton, G. S., Teanby, N. A., Irwin, P. G. J. & Bjoraker, G. L. Methane and its isotopologues on Saturn from Cassini/CIRS

observations. *Icarus* **199**, 351–367 (2009).

26. Polanski, A. S., Crossfield, I. J. M., Howard, A. W., Isaacson, H. & Rice, M. Chemical abundances for 25 JWST exoplanet host stars with KeckSpec. *Res. Not. Am. Astron. Soc.* **6**, 155 (2022).

27. Lodders, K. & Fegley, B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. I. Carbon, nitrogen, and oxygen. *Icarus* **155**, 393–424 (2002).

28. Woitke, P. et al. Equilibrium chemistry down to 100 K. Impact of silicates and phyllosilicates on the carbon to oxygen ratio. *Astron. Astrophys.* **614**, A1 (2018).

29. Gandhi, S., Madhusudhan, N., Hawker, G. & Piette, A. HyDRA-H: simultaneous hybrid retrieval of exoplanetary emission spectra. *Astron. J.* **158**, 228 (2019).

30. Line, M. R. et al. A solar C/O and sub-solar metallicity in a hot Jupiter atmosphere. *Nature* **598**, 580–584 (2021).

31. Li, C. et al. The water abundance in Jupiter's equatorial zone. *Nat. Astron.* **4**, 609–616 (2020).

32. Atreya, S. K. et al. The origin and evolution of Saturn: A post-Cassini perspective. Preprint at https://arxiv.org/abs/2205.06914 (2022).

33. Tsai, S.-M. et al. A comparative study of atmospheric chemistry with VULCAN. *Astrophys. J.* **923**, 264 (2021).

34. Agúndez, M., Parmentier, V., Venot, O., Hersant, F. & Selsis, F. Pseudo 2D chemical model of hot-Jupiter atmospheres: application to HD 209458b and HD 189733b. *Astron. Astrophys.* **564**, A73 (2014).

35. Hu, R. Photochemistry and spectral characterization of temperate and gas-rich exoplanets. *Astrophys. J.* **921**, 27 (2021).

36. Steinrueck, M. E., Parmentier, V., Showman, A. P., Lothringer, J. D. & Lupu, R. E. The effect of 3D transport-induced disequilibrium carbon

chemistry on the atmospheric structure, phase curves, and emission spectra of hot Jupiter HD 189733b. *Astrophys. J.* **880**, 14 (2019).

37. Molaverdikhani, K., Henning, T. & Mollière, P. From cold to hot irradiated gaseous exoplanets: fingerprints of chemical disequilibrium in atmospheric spectra. *Astrophys. J.* **883**, 194 (2019).

38. Molaverdikhani, K., Henning, T. & Mollière, P. The role of clouds on the depletion of methane and water dominance in the transmission spectra of irradiated exoplanets. *Astrophys. J.* **899**, 53 (2020).

39. Drummond, B. et al. Implications of three-dimensional chemical transport in hot Jupiter atmospheres: results from a consistently coupled chemistry–radiation–hydrodynamics model. *Astron. Astrophys.* **636**, A68 (2020).

40. Feinstein, A. D. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRISS. *Nature* https://doi.org/10.1038/s41586-022-05674-1 (2023).

41. Alderson, L. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H. *Nature* https://doi.org/10.1038/s41586-022-05591-3 (2023).

42. Rustamkulov, Z. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM. *Nature* https://doi.org/10.1038/s41586-022-05677-y (2023).

43. Ali-Dib, M. Disentangling hot Jupiters formation location from their chemical composition. *Mon. Not. R. Astron. Soc.* **467**, 2845–2854 (2017).

44. Cridland, A. J., van Dishoeck, E. F., Alessi, M. & Pudritz, R. E. Connecting planet formation and astrochemistry. A main sequence for C/O in hot exoplanetary atmospheres. *Astron. Astrophys.* **632**, A63 (2019).

45. Danielson, R. E. The infrared spectrum of Jupiter. *Astrophys. J.* **143**, 949 (1966).

46. Schlawin, E. et al. JWST noise floor. I. Random error sources in JWST NIRCam time series. *Astron. J.* **160**, 231 (2020).

47. Wong, M. H., Mahaffy, P. R., Atreya, S. K., Niemann, H. B. & Owen, T. C. Updated Galileo probe mass spectrometer measurements of carbon, oxygen, nitrogen, and sulfur on Jupiter. *Icarus* **171**, 153–170 (2004).

48. Karkoschka, E. & Tomasko, M. G. The haze and methane distributions on Neptune from HST-STIS spectroscopy. *Icarus* **211**, 780–797 (2011).

49. Sromovsky, L. A., Fry, P. M. & Kim, J. H. Methane on Uranus: the case for a compact CH_4 cloud layer at low latitudes and a severe CH_4 depletion at high-latitudes based on re-analysis of Voyager occultation measurements and STIS spectroscopy. *Icarus* **215**, 292–312 (2011).

50. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).

51. Bryant, E. M. et al. Simultaneous TESS and NGTS transit observations of WASP-166 b. *Mon. Not. R. Astron. Soc.* **494**, 5872–5881 (2020).

52. O'Brien, S. M. et al. Scintillation-limited photometry with the 20-cm NGTS telescopes at Paranal Observatory. *Mon. Not. R. Astron. Soc.* **509**, 6111–6118 (2022).

53. Rieke, M. J., Kelly, D. & Horner, S. in *Cryogenic Optical Systems and Instruments XI* Proceedings of SPIE Vol. 5904 (eds Heaney, J. B. & Burriesci, L. G.) 1–8 (SPIE, 2005).

54. Greene, T. P. et al. λ = 2.4 to 5 μm spectroscopy with the James Webb Space Telescope NIRCam instrument. *J. Astron. Telesc. Instrum. Syst.* **3**, 035001 (2017).

55. Bell, T. J. et al. Eureka!: an end-to-end pipeline for JWST time-series observations. *J. Open Source Softw.* **7**, 4503 (2022).

56. Horne, K. An optimal extraction algorithm for CCD spectroscopy. *Publ. Astron. Soc. Pac.* **98**, 609–617 (1986).

57. Lendl, M. et al. FORS2 observes a multi-epoch transmission spectrum of the hot Saturn-mass exoplanet WASP-49b. *Astron. Astrophys.* **587**, A67 (2016).

58. Lendl, M. et al. Signs of strong Na and K absorption in the transmission spectrum of WASP-103b. *Astron. Astrophys.* **606**, A18 (2017).

59. van Dokkum, P. G. Cosmic-ray rejection by Laplacian edge detection. *Publ. Astron. Soc. Pac.* **113**, 1420–1427 (2001).

60. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pac.* **125**, 306 (2013).

61. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* https://doi.org/10.7717/peerj-cs.55 (2016).

62. Foreman-Mackey, D. et al. exoplanet: gradient-based probabilistic inference for exoplanet data & other astronomical time series. *J. Open Source Softw.* **6**, 3285 (2021).

63. Luger, R. et al. starry: analytic occultation light curves. *Astron. J.* **157**, 64 (2019).

64. Wakeford, H. & Grant, D. Exo-TiC/ExoTiC-LD: ExoTiC-LD v2.1 zenodo release. *Zenodo* https://doi.org/10.5281/zenodo.6809899 (2022).

65. Laginja, I. & Wakeford, H. ExoTiC-ISM: a Python package for marginalised exoplanet transit parameters across a grid of systematic instrument models. *J. Open Source Softw.* **5**, 2281 (2020).

66. Laginja, I. & Wakeford, H. ExoTiC-ISM v2.0.0. *Zenodo* https://doi.org/10.5281/zenodo.3923986 (2020).

67. Morello, G., Tsiaras, A., Howarth, I. D. & Homeier, D. High-precision stellar limb-darkening in exoplanetary transits. *Astron. J.* **154**, 111 (2017).

68. Morello, G. et al. The ExoTETHyS package: tools for exoplanetary transits around host stars. *Astron. J.* **159**, 75 (2020).

69. Kipping, D. M. Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. *Mon. Not. R. Astron. Soc.* **435**, 2152–2160 (2013).

70. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for −5.0 <= log[M/H] <= +1, 2000 K <= T_{eff} <= 50000 K at several surface gravities. *Astron. Astrophys.* **363**, 1081–1190 (2000).

71. Tremblin, P. et al. Fingering convection and cloudless models for cool brown dwarf atmospheres. *Astrophys. J. Lett.* **804**, L17 (2015).

72. Goyal, J. M. et al. A library of ATMO forward model transmission spectra for hot Jupiter exoplanets. *Mon. Not. R. Astron. Soc.* **474**, 5158–5185 (2018).

73. Goyal, J. M. et al. A library of self-consistent simulated exoplanet atmospheres. *Mon. Not. R. Astron. Soc.* **498**, 4680–4704 (2020).

74. Hauschildt, P. H., Allard, F. & Baron, E. The NextGen Model Atmosphere grid for 3000 <= T_{eff} <= 10,000 K. *ApJ* **512**, 377–385 (1999).

75. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885–895 (2001).

76. Lothringer, J. D. & Barman, T. S. The PHOENIX exoplanet retrieval algorithm and using H^- opacity as a probe in ultrahot Jupiters. *Astron. J.* **159**, 289 (2020).

77. Batalha, N. E., Marley, M. S., Lewis, N. K. & Fortney, J. J. Exoplanet reflected-light spectroscopy with PICASO. *Astrophys. J.* **878**, 70 (2019).

78. Mukherjee, S., Batalha, N. E., Fortney, J. J. & Marley, M. S. PICASO 3.0: a one-dimensional climate model for giant planets and brown dwarfs. *Astrophys. J.* **942**, 71 (2023).

79. McKay, C. P., Pollack, J. B. & Courtin, R. The thermal structure of Titan's atmosphere. *Icarus* **80**, 23–53 (1989).

80. Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D. & Freedman, R. Comparative planetary atmospheres: models of TrES-1 and HD 209458b. *Astrophys. J. Lett.* **627**, L69–L72 (2005).

81. Marley, M. S. et al. The Sonora brown dwarf atmosphere and evolution models. I. Model description and application to cloudless atmospheres in rainout chemical equilibrium. *Astrophys. J.* **920**, 85 (2021).

82. Barber, R. J., Tennyson, J., Harris, G. J. & Tolchenov, R. N. A high-accuracy computed water line list. *Mon. Not. R. Astron. Soc.* **368**, 1087–1094 (2006).

83. Tennyson, J. & Yurchenko, S. The ExoMol atlas of molecular opacities. *Atoms* **6**, 26 (2018).

84. Tennyson, J. & Yurchenko, S. N. ExoMol: molecular line lists for exoplanet and other atmospheres. *Mon. Not. R. Astron. Soc.* **425**, 21–33 (2012).

85. Yurchenko, S. N., Tennyson, J., Barber, R. J. & Thiel, W. Vibrational transition moments of CH_4 from first principles. *J. Mol. Spectrosc.* **291**, 69–76 (2013).

86. Yurchenko, S. N. & Tennyson, J. ExoMol line lists—IV. The rotation–vibration spectrum of methane up to 1500 K. *Mon. Not. R. Astron. Soc.* **440**, 1649–1661 (2014).

87. Huang, X., Gamache, R. R., Freedman, R. S., Schwenke, D. W. & Lee, T. J. Reliable infrared line lists for 13 CO_2 isotopologues up to E'=18,000 cm^{-1} and 1500 K, with line shape parameters. *J. Quant. Spectrosc. Radiat. Transf.* **147**, 134–144 (2014).

88. Rothman, L. S. et al. The HITRAN2012 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **130**, 4–50 (2013).

89. Azzam, A. A. A., Lodi, L., Yurchenko, S. N. & Tennyson, J. The dipole moment surface for hydrogen sulfide H_2S. *J. Quant. Spectrosc. Radiat. Transf.* **161**, 41–49 (2015).

90. Rooney, C. M., Batalha, N. E., Gao, P. & Marley, M. S. A new sedimentation model for greater cloud diversity in giant exoplanets and brown dwarfs. *Astrophys. J.* **925**, 33 (2022).

91. Ackerman, A. S. & Marley, M. S. Precipitating condensation clouds in substellar atmospheres. *Astrophys. J.* **556**, 872–884 (2001).

92. Lothringer, J. D. et al. A new window into planet formation and migration: refractory-to-volatile elemental ratios in ultra-hot Jupiters. *Astrophys. J.* **914**, 12 (2021).

93. Drummond, B. et al. The effects of consistent chemical kinetics calculations on the pressure–temperature profiles and emission spectra of hot Jupiters. *Astron. Astrophys.* **594**, A69 (2016).

94. Kawashima, Y. & Min, M. Implementation of disequilibrium chemistry to spectral retrieval code ARCiS and application to 16 exoplanet transmission spectra. Indication of disequilibrium chemistry for HD 209458b and WASP-39b. *Astron. Astrophys.* **656**, A90 (2021).

95. Benneke, B. & Seager, S. How to distinguish between cloudy mini-Neptunes and water/volatile-dominated super-Earths. *Astrophys. J.* **778**, 153 (2013).

96. Pinhas, A. & Madhusudhan, N. On signatures of clouds in exoplanetary transit spectra. *Mon. Not. R. Astron. Soc.* **471**, 4355–4373 (2017).

97. Kitzmann, D. & Heng, K. Optical properties of potential condensates in exoplanetary atmospheres. *Mon. Not. R. Astron. Soc.* **475**, 94–107 (2018).

98. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104 (2008).

99. Kreidberg, L. batman: Basic Transit Model Calculation in Python. *Publ. Astron. Soc. Pac.* **127**, 1161 (2015).

100. JWST Calibration Pipeline Developers. jwst: Python library for science observations from the James Webb Space Telescope. *GitHub* https://github.com/spacetelescope/jwst (2022).

101. Lee, E. K. H. et al. 3D radiative transfer for exoplanet atmospheres. gCMCRT: a GPU-accelerated MCRT code. *Astrophys. J.* **929**, 180 (2022).

102. Kreidberg, L. et al. Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *Nature* **505**, 69–72 (2014).

Acknowledgements

This work is based in part on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. These observations are associated with programme #1366. Support for this programme was provided by NASA through a grant from STScI. This work is based in part on data collected under the NGTS project at the European Southern Observatory's La Silla Paranal Observatory. The NGTS facility is operated by a consortium of institutes with support from the UK Science and Technology Facilities Council (STFC) under projects ST/M001962/1, ST/S002642/1 and ST/W003163/1. This paper includes data collected by the TESS mission, obtained from MAST at STScI. Funding for the TESS mission is provided by the NASA's Science Mission Directorate. This article is supported by the UK Research and Innovation (UKRI) to fund open access through the University of Warwick.

Author information

Authors and Affiliations

1. Centre for Exoplanets and Habitability, University of Warwick, Coventry, UK

   Eva-Maria Ahrer, Peter J. Wheatley, Matthew P. Battley & David R. Anderson

2. Department of Physics, University of Warwick, Coventry, UK

   Eva-Maria Ahrer, Peter J. Wheatley, Matthew P. Battley, Samuel Gill, David R. Anderson, Daniel Bayliss & Kevin Heng

3. Johns Hopkins APL, Laurel, MD, USA

Kevin B. Stevenson, Jacob Lustig-Yaeger, E. M. May & Kristin S. Sotzen

4. Steward Observatory, University of Arizona, Tucson, AZ, USA

   Megan Mansfield & Everett Schlawin

5. Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA

   Sarah E. Moran

6. Department of Physics and Astronomy, University of Kansas, Lawrence, KS, USA

   Jonathan Brande & Ian J. M. Crossfield

7. Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain

   Giuseppe Morello & Enric Palle

8. Departamento de Astrofísica, Universidad de La Laguna (ULL), Tenerife, Spain

   Giuseppe Morello

9. INAF- Palermo Astronomical Observatory, Piazza del Parlamento, Palermo, Italy

   Giuseppe Morello

10. Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, USA

    Catriona A. Murray & Zachory K. Berta-Thompson

11. Space Telescope Science Institute, Baltimore, MD, USA

    Nikolay K. Nikolov, Néstor Espinoza & Quentin Changeat

12. Département d’Astronomie, Université de Genève, Sauverny, Switzerland

Dominique J. M. Petit dit de la Roche & Monika Lendl

13. Max Planck Institute for Astronomy, Heidelberg, Germany

Sebastian Zieba, Laura Kreidberg, Luigi Mancini, Thomas Mikal-Evans & Karan Molaverdikhani

14. Leiden Observatory, University of Leiden, Leiden, The Netherlands

Sebastian Zieba & Yamila Miguel

15. NASA Ames Research Center, Moffett Field, CA, USA

Natasha E. Batalha

16. Jet Propulsion Laboratory, Pasadena, CA, USA

Mario Damiano

17. School of Earth and Planetary Sciences (SEPS), National Institute of Science Education and Research (NISER), HBNI, Jatani, India

Jayesh M. Goyal

18. Department of Physics, Utah Valley University, Orem, UT, USA

Joshua D. Lothringer

19. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

Sagnick Mukherjee, Kazumasa Ohno, Natalie M. Batalha, Aarynn L. Carter & Jonathan J. Fortney

20. Astrobiology Program, University of California, Santa Cruz, Santa Cruz, CA, USA

Natalie M. Batalha

21. Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA

Jacob L. Bean & Adina D. Feinstein

22. Department of Astronomy, University of Wisconsin-Madison, Madison, WI, USA

Thomas G. Beatty

23. Department of Physics and Institute for Research on Exoplanets, Université de Montréal, Montreal, Quebec, Canada

Björn Benneke, Caroline Piaulet, Pierre-Alexis Roy & Jake Taylor

24. Space Research Institute, Austrian Academy of Sciences, Graz, Austria

Patricio E. Cubillos & Ludmila Carone

25. INAF – Osservatorio Astrofisico di Torino, Pino Torinese, Italy

Patricio E. Cubillos

26. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

Tansu Daylan

27. Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA

Néstor Espinoza, David K. Sing & Natalie H. Allen

28. Earth and Planets Laboratory, Carnegie Institution for Science, Washington DC, USA

Peter Gao & Munazza K. Alam

29. School of Physics, Trinity College Dublin, Dublin, Ireland

Neale P. Gibson

30. Planetary Sciences Group, Department of Physics and Florida Space Institute, University of Central Florida, Orlando, FL, USA

Joseph Harrington

31. Astrophysics Section, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Renyu Hu

32. Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

Renyu Hu & Heather A. Knutson

33. Department of Astronomy and Carl Sagan Institute, Cornell University, Ithaca, NY, USA

Nikole K. Lewis, Laura Flagg, Ryan J. MacDonald & Jake D. Turner

34. School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

Michael R. Line, Luis Welbanks & Evgenya L. Shkolnik

35. Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA

Mercedes López-Morales, Diana K. Powell & James Kirk

36. Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

Vivien Parmentier, Shang-Min Tsai & Jake Taylor

37. Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Nice, France

Vivien Parmentier

38. Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA

David K. Sing, Zafar Rustamkulov & Kristin S. Sotzen

39. School of Physics, University of Bristol, Bristol, UK

Hannah R. Wakeford & Lili Alderson

40. School of Physical Sciences, The Open University, Milton Keynes, UK

Joanna K. Barstow

41. BAER Institute, NASA Ames Research Center, Moffet Field, CA, USA

Taylor J. Bell

42. Department of Physics, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Jasmina Blecic

43. Center for Astro, Particle and Planetary Physics (CAP3), New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Jasmina Blecic

44. Mullard Space Science Laboratory, University College London, Dorking, UK

Edward M. Bryant

45. School of Physics and Astronomy, University of Leicester, Leicester, UK

Matthew R. Burleigh & S. L. Casewell

46. European Space Agency, Space Telescope Science Institute, Baltimore, MD, USA

Quentin Changeat & Sarah Kendrew

47. Department of Physics and Astronomy, University College London, London, UK

Quentin Changeat

48. Centre for Exoplanet Science, University of St Andrews, St Andrews, UK

Katy L. Chubb

49. Leiden Observatory, Leiden University, Leiden, The Netherlands

Nicolas Crouzet

50. Institute of Astronomy, Department of Physics and Astronomy, KU Leuven, Leuven, Belgium

Leen Decin

51. Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, The Netherlands

Jean-Michel Désert

52. Department of Physics and Astronomy, University of Delaware, Newark, DE, USA

John E. Gizis

53. University Observatory Munich, Ludwig Maximilian University, Munich, Germany

Kevin Heng & Karan Molaverdikhani

54. ARTORG Center for Biomedical Engineering, University of Bern, Bern, Switzerland

Kevin Heng

55. Institute for Astrophysics, University of Vienna, Vienna, Austria

Nicolas Iro

56. Department of Astronomy, University of Maryland, College Park, MD, USA

Eliza M.-R. Kempton & Thaddeus D. Komacek

57. Department of Physics, Imperial College London, London, UK

James Kirk

58. Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, Gif-sur-Yvette, France

Pierre-Olivier Lagage

59. Laboratoire d'Astrophysique de Bordeaux, Université de Bordeaux, Pessac, France

Jérémy Leconte

60. Department of Astronomy, University of Michigan, Ann Arbor, MI, USA

Ryan J. MacDonald

61. Department of Physics, University of Rome “Tor Vergata”, Rome, Italy

Luigi Mancini

62. INAF - Turin Astrophysical Observatory, Pino Torinese, Italy

Luigi Mancini

63. Department of Physics and Astronomy, University of Exeter, Exeter, UK

N. J. Mayne

64. SRON Netherlands Institute for Space Research, Leiden, The Netherlands

Yamila Miguel

65. Exzellenzcluster Origins, Garching, Germany

Karan Molaverdikhani

66. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Benjamin V. Rackham

67. Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Benjamin V. Rackham

68. Astronomy Department and Van Vleck Observatory, Wesleyan University, Middletown, CT, USA

Seth Redfield

69. Institute of Astronomy, University of Cambridge, Cambridge, UK

Laura K. Rogers

70. Maison de la Simulation, CEA, CNRS, Univ. Paris-Sud, UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

P. Tremblin

71. Department of Physics, Brown University, Providence, RI, USA

Gregory S. Tucker

72. Planetary Science Institute, Tucson, AZ, USA

Miguel de Val-Borro

73. Université de Paris Cité and Univ Paris Est Creteil, CNRS, LISA, Paris, France

Olivia Venot

74. Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA

Xi Zhang

Contributions

All authors played a significant role in one or more of the following: development of the original proposal, management of the project, definition of the target list and observation plan, analysis of the data, theoretical modelling and preparation of this manuscript. Some specific contributions are listed as follows. N.M.B., J.L.B. and K.B.S. provided overall programme leadership and management. E.-M.A. and K.B.S. led the efforts for this manuscript. D.K.S., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., K.B.S., L.K., M.L.-M., M.R.L., N.M.B., V.P. and Z.K.B.-T. made significant contributions to the design of the programme. K.B.S. generated the observing plan with input from the team. E.S., N.E. and T.G.B. provided instrument expertise. B.B., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., L.K.,

M.L.-M., M.R.L., N.M.B. and Z.K.B.-T. led or co-led working groups and/or contributed to significant strategic planning efforts like the design and implementation of the pre-launch Data Challenges. A.L.C., D.K.S., E.S., N.E., N.P.G., T.G.B. and V.P. generated simulated data for pre-launch testing of methods. C.A.M., D.J.M.P., E.-M.A., E.S., J. Brande, K.B.S., M.D., M.M., N.K.N. and S.Z. reduced the data, modelled the light curves and produced the planetary spectrum. J.M.G., J.L., K.O., N.E.B., S.E.M. and S.M. generated theoretical model grids for comparison with data. C.A.M., D.J.M.P., E.-M.A., E.S., J.L., K.B.S., M.M., N.K.N., P.J.W. and S.E.M. made significant contributions to the writing of this manuscript. C.A.M., E.-M.A., K.B.S., M.P.B., M.M., P.J.W., S.G., S.E.M. and Z.K.B.-T. generated figures for this paper. B.V.R., E.L.S., G.M., J. Brande, S.Z. and Z.K.B.-T. contributed to the writing of this manuscript. J.H., M.R.L., N.K.L., P.G., P.J.W., R.H., T.D. and T.G.B. provided significant feedback to the manuscript.

Corresponding authors

Correspondence to Eva-Maria Ahrer or Kevin B. Stevenson.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Photometric monitoring of WASP-39 (top) and individual transit observations (bottom) using NGTS (magenta) and TESS (dark purple).

The black marks indicate the times of the four JWST ERS transit observations. The monitoring light curve shows evidence for optical variability, but with an RMS amplitude of only 0.06% in NGTS. The times of the individual transit observations are indicated on the top panel, and they are all consistent with transits free of starspot crossings or other features associated with stellar activity.

Extended Data Fig. 2 Raw NIRCam image of the LW (top) and SW (bottom) channels.

The faint horizontal stripes seen in the LW channel originate from neighbouring objects. The SW channel is able to track changes in alignment for individual mirror segments. No impactful tilt events were noted in this observation.

Extended Data Fig. 3 Median NIRCam frame, after curvature correction and background subtraction, shown as the full 2D frame (left) and a vertical slice (right).

Left: curvature-corrected, background-subtracted, median frame. We perform optimal spectral extraction on the pixels in between the green dashed lines. We use the pixels outside of the two orange solid lines for background subtraction. The flux spans −200–1000 electrons, thus drawing attention to the residual background features. Right: vertical slice depicting the flux averaged over detector pixels 855 to 865. The background region clearly demonstrates some low-level residual structure.

Extended Data Fig. 4 Normalized root-mean-square error as a function of bin size for all spectroscopic channels.

The red line shows the expected relationship for perfect Gaussian white noise. The black lines show the observed noise from each spectroscopic channel for the Eureka! long-wavelength reduction. Values for all channels are normalized by dividing by the value for a bin size of 1 in order to compare bins with different noise levels. The black lines closely follow the red line out to large bin sizes of ≈30 (≈0.5-h timescales), which demonstrates that the residuals to the fit are dominated by white Gaussian noise.

Extended Data Fig. 5 The transit spectrum of WSP-39b as determined by our independent analysis using JWST's NIRCam instrument (top) and the respective differences between our results (bottom).

Top: transmission spectra from our reductions when including additional data on the blue and red edges (now spanning 2.405–4.055 μm). This demonstrates the large error bars and diverging data points near the edges of the NIRCam bandpass in the LW spectroscopic channel. Bottom: the differences in retrieved transmission spectra by subtracting the Eureka! spectrum from the other three reduced spectra shown in the top panel. This shows the strong agreement between the spectra; however, we do note minor disagreements at shorter wavelengths that we attribute to differences in the treatment of limb-darkening effects within the individual fitting methods.

Extended Data Fig. 6 JWST's NIRCam data of WASP-39 as a function of time and wavelength for each independent reduction (top) and their residuals (bottom).

Top: time-series NIRCam data for the WASP-39b system, from three independent spectral extractions. Colour represents relative brightness at each time and wavelength, normalized by the median stellar spectrum. Bottom: resulting residuals after fitting the time-series NIRCam data.

Extended Data Fig. 7 Measured transmission spectrum compared to atmospheric forward model grids.

Top: the single best fit for each model grid (shown as solid coloured lines; PICASO 3.0, ATMO, PHOENIX), fits the planet spectrum (Eureka! reduction) with $\chi_{\nu}^{2} \le 1.22$. All single best fits prefer at least solar metallicity and substantial cloud cover. Also shown as a grey dashed line is a solar metallicity, stellar C/O ratio atmospheric model, demonstrating the lack of methane absorption seen in the spectrum. Because we can put an upper limit on the CH_4 abundance, the preferred C/O ratio found by the model grids is substellar. Bottom: residuals of each best fit, shown as the model spectrum subtracted from the reduced spectrum and divided by the uncertainty in transit depth. The residuals show wavelength-dependent correlations, the origin of which are unknown and left for a future study.

Extended Data Fig. 8 Our JWST/NIRCam spectrum compared with existing HST/WFC3 data.

As in Extended Data Fig. 7, but with the addition of HST/WFC3 data from 0.8 to 1.65 µm, showing the comparable precision and complementary wavelength coverage offered by the combination of NIRCam and HST/WFC3.

Extended Data Fig. 9 Gaussian residual fitting of H_2O and CO_2.

The blue points show the residual features left after subtracting out the gas in question (CO_2, top, and H_2O, bottom) from the single best-fit model. The Gaussian model ensemble fit to the residual is shown in red; the best-fit Gaussian ensemble to a flat-line model is shown in blue. We strongly detect H_2O at nearly 16σ and show weak evidence for CO_2 (small feature at 2.6 µm) at 1.9σ.

Extended Data Table 1 Details of the two methods used to fit the short-wavelength (SW) photometry
Extended Data Table 2 Details of the four fitting methods used to fit the long-wavelength (LW) spectroscopy
Extended Data Table 3 Best-fit orbital parameters from both short-wavelength (SW) and white-light long-wavelength (LW) fits

Source data

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Ahrer, EM., Stevenson, K.B., Mansfield, M. *et al.* Early Release Science of the exoplanet WASP-39b with JWST NIRCam. *Nature* **614**, 653–658 (2023). https://doi.org/10.1038/s41586-022-05590-4

- Received: 16 September 2022
- Accepted: 24 November 2022
- Published: 09 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05590-4

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H**

  - Lili Alderson
  - Hannah R. Wakeford
  - Xi Zhang

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM**

    - Z. Rustamkulov
    - D. K. Sing
    - S. Zieba

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRISS**

    - Adina D. Feinstein
    - Michael Radica
    - Xi Zhang

  Nature (2023)

JWST opens a window on exoplanet skies

- Julia V. Seidel
- Louise D. Nielsen
- Subhajit Sarkar

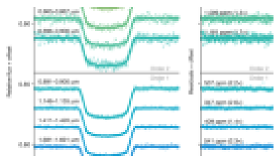
Nature News & Views Forum 15 Feb 2023

A population of red candidate massive galaxies ~600 Myr after the Big Bang

- Ivo Labbé
- Pieter van Dokkum
- Bingjie Wang

Nature Article 22 Feb 2023

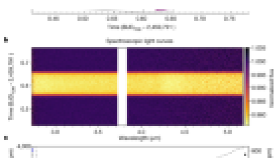
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05590-4



Early Release Science of the exoplanet WASP-39b with JWST NIRISS

09 January 2023

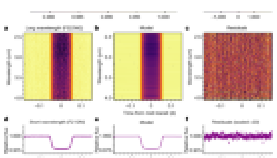
Adina D. Feinstein, Michael Radica, … Xi Zhang



Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H

09 January 2023

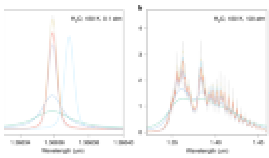
Lili Alderson, Hannah R. Wakeford, … Xi Zhang



Early Release Science of the exoplanet WASP-39b with JWST NIRCam

09 January 2023

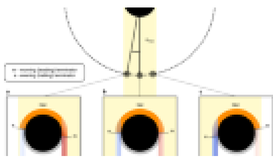
Eva-Maria Ahrer, Kevin B. Stevenson, … Xi Zhang



The impending opacity challenge in exoplanet atmospheric characterization

15 September 2022

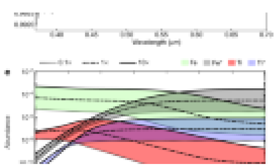
Prajwal Niraula, Julien de Wit, … Roman V. Kochanov



Titanium oxide and chemical inhomogeneity in the atmosphere of the exoplanet WASP-189 b

27 January 2022

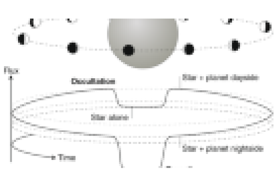
Bibiana Prinoth, H. Jens Hoeijmakers, … Kevin Heng



Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b

15 August 2018

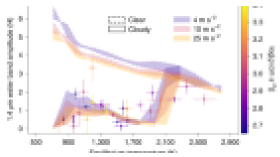
H. Jens Hoeijmakers, David Ehrenreich, … Luca Di Fabrizio



Highlights of exoplanetary science from Spitzer

14 May 2020

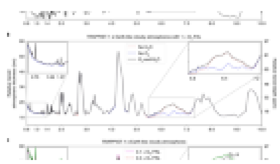
Drake Deming & Heather A. Knutson



Aerosol composition of hot giant exoplanets dominated by silicates and hydrocarbon hazes

25 May 2020

Peter Gao, Daniel P. Thorngren, … Xi Zhang



Sensitive probing of exoplanetary oxygen via mid-infrared collisional absorption

06 January 2020

Thomas J. Fauchez, Geronimo L. Villanueva, … Shawn D. Domagal-Goldman

- Article
- Open Access
- Published: 09 January 2023

Early Release Science of the exoplanet WASP-39b with JWST

NIRSpec PRISM

- Z. Rustamkulov ,
- D. K. Sing,
- S. Mukherjee,
- E. M. May,
- J. Kirk,
- E. Schlawin,
- M. R. Line,
- C. Piaulet,
- A. L. Carter,
- N. E. Batalha,
- J. M. Goyal,
- M. López-Morales,
- J. D. Lothringer,
- R. J. MacDonald,
- S. E. Moran,
- K. B. Stevenson,
- H. R. Wakeford,
- N. Espinoza,
- J. L. Bean,
- N. M. Batalha,
- B. Benneke,
- Z. K. Berta-Thompson,
- I. J. M. Crossfield,
- P. Gao,
- L. Kreidberg,
- D. K. Powell,
- P. E. Cubillos,
- N. P. Gibson,
- J. Leconte,
- K. Molaverdikhani,
- N. K. Nikolov,
- V. Parmentier,
- P. Roy,
- J. Taylor,

- J. D. Turner,
- P. J. Wheatley,
- K. Aggarwal,
- E. Ahrer,
- M. K. Alam,
- L. Alderson,
- N. H. Allen,
- A. Banerjee,
- S. Barat,
- D. Barrado,
- J. K. Barstow,
- T. J. Bell,
- J. Blecic,
- J. Brande,
- S. Casewell,
- Q. Changeat,
- K. L. Chubb,
- N. Crouzet,
- T. Daylan,
- L. Decin,
- J. Désert,
- T. Mikal-Evans,
- A. D. Feinstein,
- L. Flagg,
- J. J. Fortney,
- J. Harrington,
- K. Heng,
- Y. Hong,
- R. Hu,
- N. Iro,
- T. Kataria,
- E. M.-R. Kempton,
- J. Krick,
- M. Lendl,
- J. Lillo-Box,
- A. Louca,
- J. Lustig-Yaeger,

- L. Mancini,
- M. Mansfield,
- N. J. Mayne,
- Y. Miguel,
- G. Morello,
- K. Ohno,
- E. Palle,
- D. J. M. Petit dit de la Roche,
- B. V. Rackham,
- M. Radica,
- L. Ramos-Rosado,
- S. Redfield,
- L. K. Rogers,
- E. L. Shkolnik,
- J. Southworth,
- J. Teske,
- P. Tremblin,
- G. S. Tucker,
- O. Venot,
- W. C. Waalkes,
- L. Welbanks,
- X. Zhang &
- …
- S. Zieba

Show authors

Nature **volume 614**, pages 659–663 (2023)

- 2666 Accesses
- 4 Citations
- 18 Altmetric
- Metrics details

Abstract

Transmission spectroscopy[1,2,3] of exoplanets has revealed signatures of water vapour, aerosols and alkali metals in a few dozen exoplanet atmospheres[4,5]. However, these previous inferences with the Hubble and Spitzer Space Telescopes were hindered by the observations' relatively narrow wavelength range and spectral resolving power, which precluded the unambiguous identification of other chemical species—in particular the primary carbon-bearing molecules[6,7]. Here we report a broad-wavelength 0.5–5.5 µm atmospheric transmission spectrum of WASP-39b[8], a 1,200 K, roughly Saturn-mass, Jupiter-radius exoplanet, measured with the JWST NIRSpec's PRISM mode[9] as part of the JWST Transiting Exoplanet Community Early Release Science Team Program[10,11,12]. We robustly detect several chemical species at high significance, including Na (19σ), H_2O (33σ), CO_2 (28σ) and CO (7σ). The non-detection of CH_4, combined with a strong CO_2 feature, favours atmospheric models with a super-solar atmospheric metallicity. An unanticipated absorption feature at 4 µm is best explained by SO_2 (2.7σ), which could be a tracer of atmospheric photochemistry. These observations demonstrate JWST's sensitivity to a rich diversity of exoplanet compositions and chemical processes.

Main

We observed one transit of WASP-39b on 10 July 2022 with JWST's Near InfraRed Spectrograph (NIRSpec)[9,13], using the PRISM mode, as part of the JWST Transiting Exoplanet Community Early Release Science Program (ERS Program 1366) (PIs: Natalie Batalha, Jacob Bean, Kevin Stevenson)[10,11]. These observations cover the 0.5–5.5 µm wavelength range at a native resolving power of 20–300. WASP-39b was selected for this JWST-ERS Program because of previous space- and ground-based observations revealing strong alkali metal absorption and several prominent H_2O bands[4,6,14,15,16], suggesting a strong signal-to-noise ratio could be obtained with JWST. However, the limited wavelength range of existing transmission spectra (0.3–1.65 µm, combined with two wide photometric Spitzer channels at 3.6 and 4.5 µm) left several important questions unresolved. Previous estimates of WASP-39b's atmospheric metallicity—a measure of the relative abundance of all gases heavier than hydrogen or

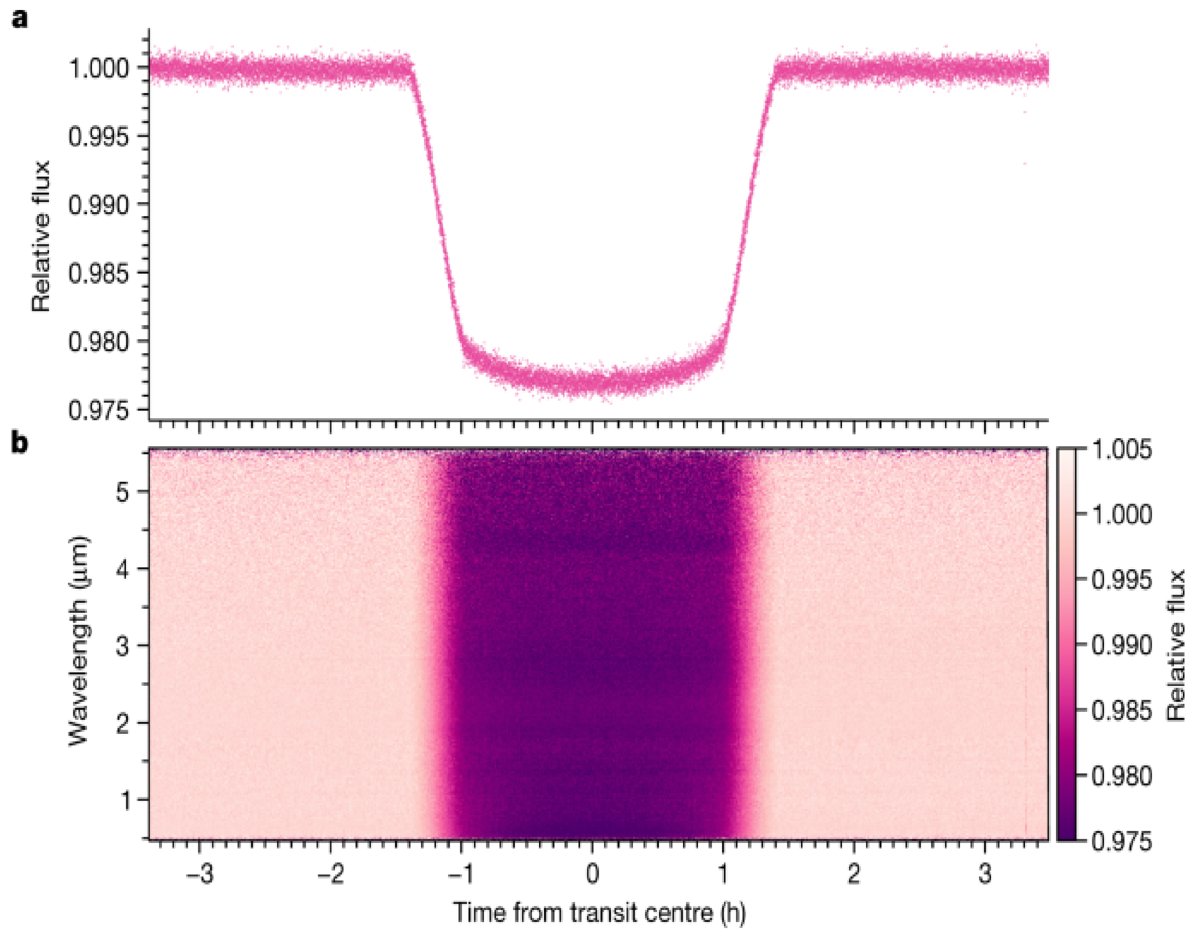
helium—vary by four orders of magnitude[6,16,17,18,19,20]. Accurate determinations of metallicity can explain formation pathways and provide greater insight into the planet's history[21]. The JWST NIRSpec PRISM observations we present here offer a more detailed view into WASP-39b's atmospheric composition than has previously been possible (see ref. [21] for an initial infrared analysis of these data).

We obtained time-series spectroscopy over 8.23 h centred around the transit event to extract the wavelength-dependent absorption by the planet's atmosphere—that is, the transmission spectrum, which probes the planet's day-night terminator region near millibar pressures. We used NIRSpec PRISM in Bright Object Time Series (BOTS) mode. WASP-39 is a bright, nearby, relatively inactive[22] G7 type star with an effective temperature of 5,400 K (ref. [8]). WASP-39's J-band magnitude of 10.66 puts it near PRISM's saturation limit, which allows us to test the effects of saturation on the quality of the resulting science compared to past measurements (Methods).

In our baseline reduction using Fast InfraRed Exoplanet Fitting for Lightcurves (FIREFLy)[23], we perform calibrations on the raw data using the jwst Python pipeline[12] and then identify and correct for bad pixels and cosmic rays. We mitigate the $1/f$ noise[9] at the group level rather than the integration level to ensure accurate slope fitting, which we find to be a crucial step for NIRSpec PRISM observations with few groups per integration.

We bin the resulting spectrophotometry in wavelength to create 207 variable-width spectral channels with roughly equal counts in each. Figure 1 shows the FIREFLy white and spectrophotometric light curves at this step in the top panel. Several absorption features are visible by-eye as darker horizontal stripes within the transit region in the two-dimensional (2D) light curve (Fig. 1), demonstrating the high quality of the raw spectrophotometry achieved by the PRISM observing mode.

Fig. 1: The light curve of WASP-39b observed by JWST NIRSpec PRISM.

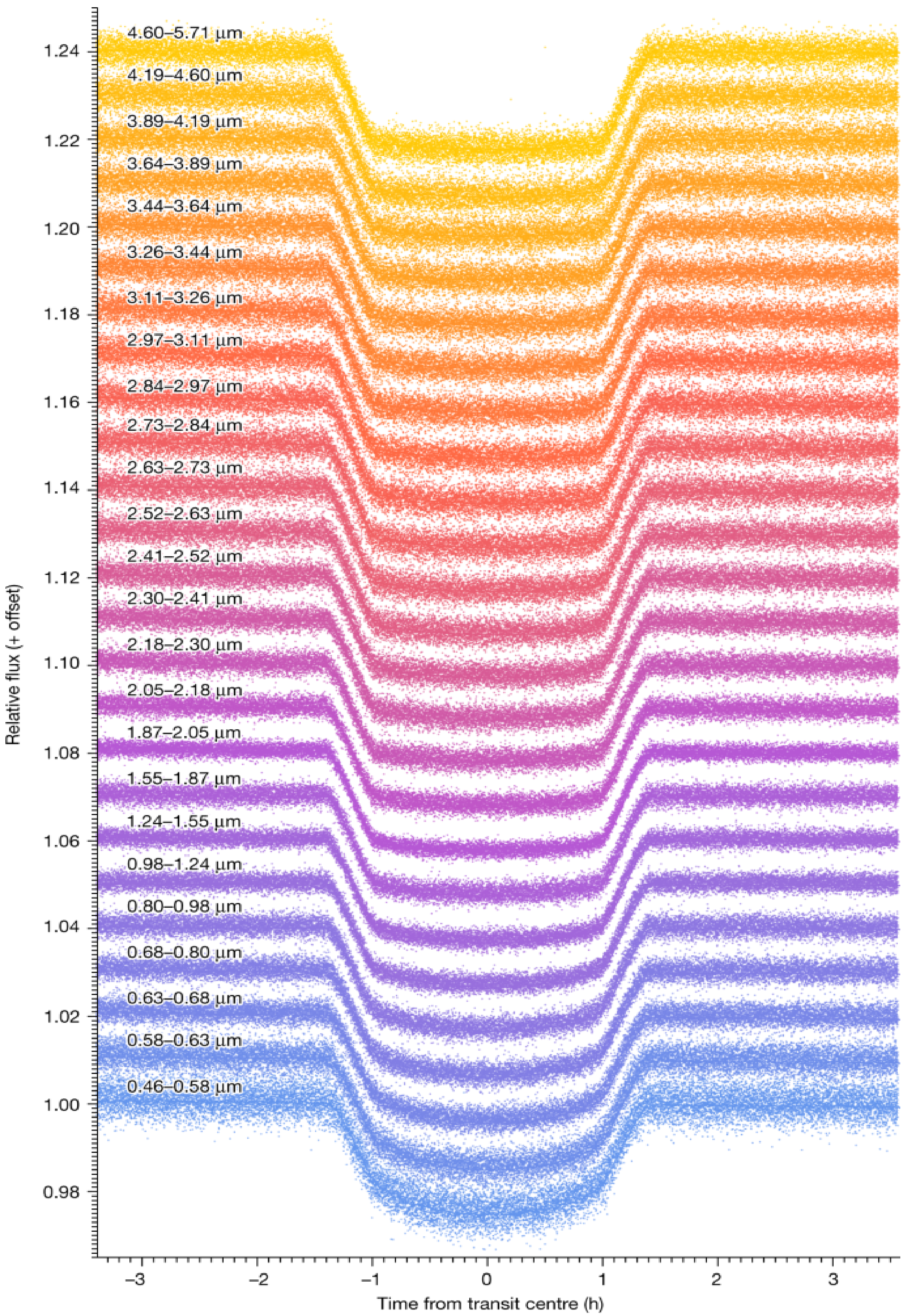


a, The normalized white light curve created by integrating over all wavelengths using the FIREFLy reduction. **b**, The binned time series (with 30 integrations per time bin) of the relative flux for each wavelength. A constant 200 ppm per hour linear trend through time has been removed from the white light curve and each spectral channel for visual clarity.

To extract the atmosphere's transmission spectrum, we fit the planet's transit depth in each wavelength bin using a limb darkened transit light curve model using the Python-based Levenburg–Marquardt least-squares algorithm lmfit[24]. The light curves show a typical photometric scatter of 0.2–1.2% per integration (1.36 s each), and the typical transit depth uncertainties vary between 50 and 200 parts per million (ppm), which is in line with near-photon-limited precision (Methods). Although we successfully measure fluxes in the saturated regions (0.8–2.3 µm), because of the lower number of groups used per integration here (1–3) the measured count rates may be adversely affected. We do not find excess red noise in the

saturated channels themselves, however, we notice large point-to-point scatter in the transit depths, which required wider wavelength binning to better match previous Hubble Space Telescope (HST) observations. Figure 2 highlights representative transit light curves spanning the entire wavelength range. These data are binned into wider wavelength channels than those used for the final transmission spectrum for ease of presentation. Light curve systematics have not been removed from these data, demonstrating the unprecedented stability and precision of the PRISM observing mode.

Fig. 2: Normalized spectrophotometric light curves for the JWST-PRISM transit of WASP-39b.

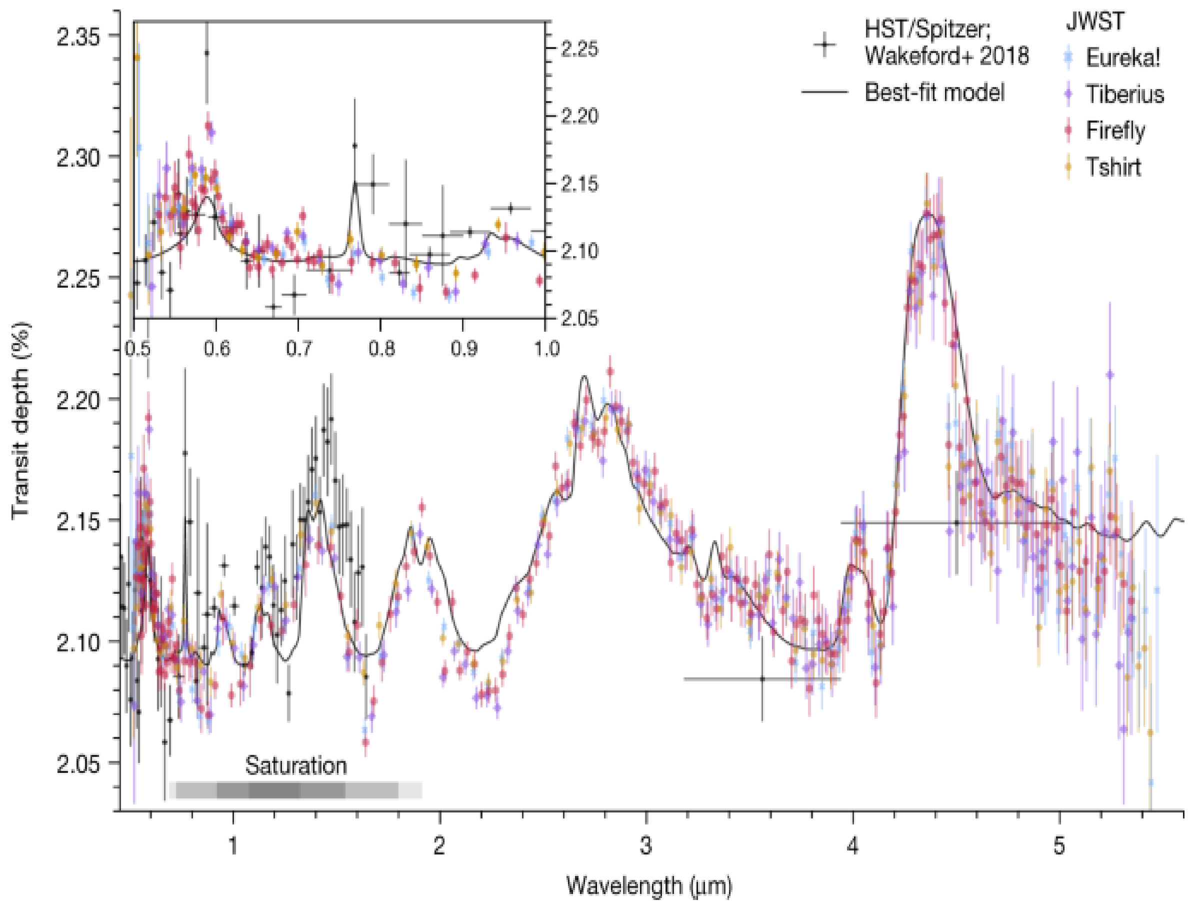


4.60–5.71 μm
4.19–4.60 μm
3.89–4.19 μm
3.64–3.89 μm
3.44–3.64 μm
3.26–3.44 μm
3.11–3.26 μm
2.97–3.11 μm
2.84–2.97 μm
2.73–2.84 μm
2.63–2.73 μm
2.52–2.63 μm
2.41–2.52 μm
2.30–2.41 μm
2.18–2.30 μm
2.05–2.18 μm
1.87–2.05 μm
1.55–1.87 μm
1.24–1.55 μm
0.98–1.24 μm
0.80–0.98 μm
0.68–0.80 μm
0.63–0.68 μm
0.58–0.63 μm
0.46–0.58 μm
Relative flux (+ offset)
1.24
1.22
1.20
1.18
1.16
1.14
1.12
1.10
1.08
1.06
1.04
1.02
1.00
0.98
–3
–2
–1
0
1
2
3
Time from transit centre (h)

The light curves were created by summing over wide wavelength channels (wavelength ranges indicated on the plot). Overplotted on each light curve are their best-fit models, which include a transit model and detector systematics. Light curve systematics have not been removed from the data.

We also compared the results from the FIREFLy reduction to three other independent reductions that use different treatments for the saturated region of the detector, limb darkening and various detector systematics (Methods). All four reductions obtain consistent results. Figure 3 shows a comparison of the four reductions. The consistency provides confidence in the accuracy of derived atmospheric parameters, demonstrating that any residual systematics are minimal and do not strongly bias results for NIRSpec PRISM observations. The transmission spectrum also agrees well with previous measurements from ground-based telescopes[15,16] as well as HST and Spitzer[6] within error (Fig. 3), indicating that we can reliably recover a spectrum at these levels of saturation. These PRISM observations offer high-quality data from 0.5–5.5 µm, with minimal contributions from systematics and at precisions generally near the photon limit (Methods). Although recovery of the saturated region (0.9–1.5 µm) is possible, caution is warranted when interpreting this portion of the spectrum (Methods). Future PRISM observations of similarly bright targets should therefore carefully consider whether saturating the spectrum is an appropriate choice for a given planet, or whether building the wavelength coverage from several transits with different complementary modes is preferable.

Fig. 3: WASP-39b transmission spectral measurements.

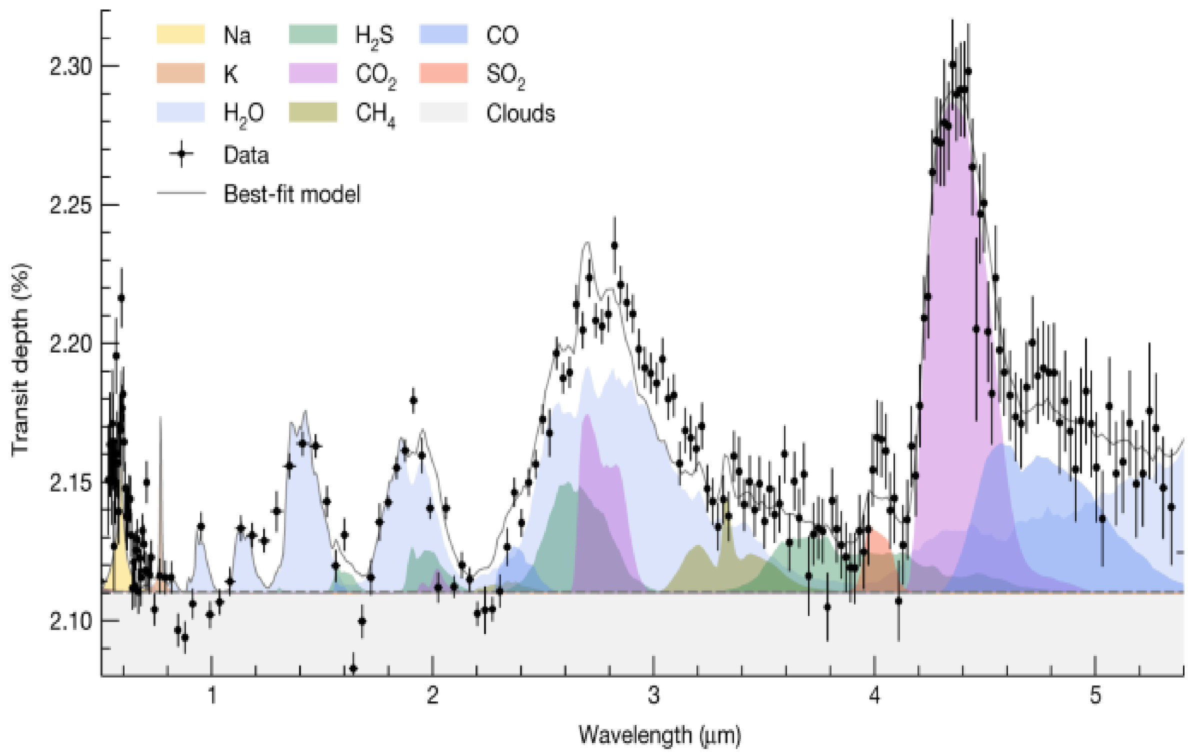


A comparison of the JWST transmission spectra obtained from the four independent reductions considered in this work (coloured points), which are all in broad agreement. Previous measurements from HST, VLT and Spitzer[6] are also shown (grey) along with our fiducial best-fit spectrum model from the PICASO 3.0 grid (black line). All the transmission spectral data have $1-\sigma$ error bars shown. The saturated region of the detector is indicated (grey bar) with the shading representative of the level of saturation (also Extended Data Fig. 6). Different reductions are presented on slightly different wavelength grids for visual purposes, the original resolution each reduction used is discussed in the Methods.

The transmission spectrum of WASP-39b from the FIREFLy reduction is shown in Fig. 4. We select the FIREFLy reduction to be our baseline reduction, but comparable results are achieved with the three other reductions presented in this work (Methods). We interpret the spectrum with grids of one-dimensional (1D) radiative–convective–thermochemical equilibrium (RCTE) models (postprocessed with some more gases

(Methods)), with a representative best-fitting model transmission spectrum shown in Fig. 4, along with opacity contributions from atoms, molecules and grey clouds. We detect the presence of H_2O by means of four pronounced independent bands (33σ, 1–2.2 µm), a prominent CO_2 feature at 4.3 µm (28σ), Na at 0.58 µm (19σ), a CO absorption band at 4.7 µm (7σ) and a grey cloud (21σ). We do not observe any significant CH_4 absorption (expected at 3.3 µm), despite predictions of its presence for atmospheres at approximately solar metallicity and place a 3σ upper limit of 5×10^{-6} on the CH_4 volume mixing ratio between 0.1 and 2 mbar. We also observe a relatively narrow absorption feature at 4.05 µm (roughy 2.7σ), which we attribute to SO_2—a potential tracer for photochemistry[25,26,27]—after an extensive search across many possible opacity sources (Methods). Using a Bayesian approach described in the Methods section, we calculate that the volume mixing ratio of SO_2 needed to explain this feature is 10^{-5}. The potential SO_2 feature is also observed at higher resolutions with JWST NIRSpec G395H (ref. [28]), adding confidence that the feature first reported as an unknown absorber[29] is a genuine feature of the planet's atmosphere. With Na detected in the atmosphere, the alkali metal, K, is also expected at optical wavelengths[14] although not detected. However, the resolution covering the narrow K absorption doublet in the optical is low, which may be preventing detection. This might also be because of detector saturation in the wavelength range where K absorption is expected. We also do not detect the presence of H_2S in the atmosphere. We note that although the best-fitting models shown in Figs. 3 and 4 have some CH_4, H_2S and K signatures, these species are not favoured by the data to the level of a detection. We determine the single best-fitting atmospheric metallicity, C/O ratio and grey cloud opacity to be ten times solar, 0.7 and $\kappa_{cld} = 10^{-2.07}$ cm^2 g^{-1}, respectively. A detailed discussion on these best-fitting parameters is presented in the Methods section.

Fig. 4: The JWST-PRISM transmission spectrum of WASP-39b with key contributions to the atmospheric spectrum.



The black points with error bars correspond to the measured FIREFLy transit depths of the spectrophotometric light curves at different wavelengths. The best-fitting model spectrum from the PICASO 3.0 grid is shown as the grey line and the coloured regions correspond to the chemical opacity contributions at specific wavelengths. The best-fitting 1D RCTE model corresponds to a super-solar metallicity and super-solar carbon-to-oxygen ratio with moderate cloud opacity (Methods). The PRISM transmission spectrum is explained by contributions from Na (19σ), H_2O (33σ), CO_2 (28σ), CO (7σ), SO_2 (2.7σ) and clouds (21σ). The data do not provide evidence of CH_4, H_2S and K absorption (Methods). Also, note that the detector was saturated to varying degrees between 0.8 and 1.9 μm. As before, the error bars are $1-\sigma$ standard deviations.

JWST/NIRSpec PRISM's power to constrain many chemical species in hot giant planet atmospheres provides new windows into their compositions and chemical processes, as we show here with WASP-39b. Using our model grids, we find that WASP-39b's best-fitting atmospheric metallicity is roughly ten times solar. In the limit of equilibrium chemistry, our non-detection of CH_4 at 3.4 μm paired with the prominence of the large CO_2 feature at 4.4 μm are indicative of a super-solar atmospheric metallicity, as

illustrated in Extended Data Fig. 9. This may point to WASP-39b's puffy envelope bearing more compositional similarity to the similarly massed ice giants than the gas giants. Moreover, the probably detection of SO_2, and its unexpectedly high estimated abundance, suggests that photochemical processes are pushing this species out of equilibrium. Photochemistry models show that sulfur compounds such as H_2S efficiently photodissociate and recombine to form SO_2 with roughly 1 ppm abundances and at 1–100 mbar pressures[26]—roughly the same pressure range probed by our transmission spectroscopy (Extended Data Fig. 10). The abundance measurement of SO_2 can therefore serve as an important tracer of the thermochemical properties of highly irradiated stratospheres and the efficiency of photochemistry. Furthermore, our detection of a qualitatively significant wavelength dependence to the planet's central transit time (Extended Data Fig. 3) suggests that these observations are sensitive to differences in the atmospheric composition at the planet's leading and trailing hemispheres. The measured roughly 20 s amplitude of this effect is in line with model expectations[30]. This indicates that such observations will be informative in exploring the three-dimensional (3D) nature of hot Jupiter atmospheres, which may give a more holistic understanding of their heat redistribution and nightside chemistry.

Methods

Data reduction

One transit of WASP-39b was observed with the NIRSpec PRISM mode, with the 8.23-h observation roughly centred around the transit event. We used NIRSpec's Bright Object Time Series (BOTS) mode with the NRSRAPID readout pattern, the S1600A1 slit (1.6" × 1.6") and the SUB512 subarray. Throughout the exposure, we recorded 21,500 integrations, each with five 0.28-s groups up the ramp. We achieved a duty cycle of 82%.

We extracted transmission spectra of WASP-39b using four different reductions with the FIREFLy, tshirt, Eureka!+ExoTEP and Tiberius pipelines. The results from all reductions are broadly consistent (Fig. 3 and Extended Data Fig. 1). We used the FIREFLy reduction as our baseline for

comparison to models throughout this paper, however, equivalent overall results can be deduced from the other reductions. Some key attributes of the reductions are compared in Extended Data Table 2. All reductions correct for 1/*f* noise: correlated frequency-dependent read noise in the images caused by detector readout and current biases in the electronics[31]. We note that as the GAINSCALE step of the JWST pipeline applies a gain correction to the raw count rate files, the counts and count rates quoted herein are in units of electrons and electrons per second, respectively.

We find that recovery of the saturated region was possible by applying several custom steps described here. Without these steps, the heavily saturated region showed a large and unexpected point-to-point scatter of several thousand ppm in the transmission spectra. We note that there was limited on-sky NIRSpec calibration data available when the data were obtained and reduced, including an incomplete detector bias image whose values were all set to zero. We used a custom bias frame for this step (private communication, S. Birkmann). Although the transmission spectra longwards of about 2 μm could be extracted without the use of this calibration, we found that bias correction was critical to extract the spectrum in the saturated region.

In addition, to recover the saturated region it was necessary to perform a reference pixel correction something that was skipped by the default jwst pipeline for NIRSpec PRISM because no official reference pixels are present in the subarray (tshirt section below). All reductions also expand the saturation flags along entire columns and only use the groups before saturation for slope fitting in these regions. With these steps, the spectra broadly matched previous HST and Very Large Telescope (VLT) observations[6], with improvement in the region with only one or two groups before saturation. We expect that as updated NIRSpec calibration data become available, the recovery of saturated regions in PRISM observations may become easier, however, we still suggest avoiding rapid saturation with less than two groups before saturation if possible, especially if that region of the spectrum is important to one's science case.

FIREFLy

We performed custom calibrations on the uncalibrated data, including $1/f$ noise destriping[9] at the group level, bad and hot pixel cleaning, cosmic ray removal and 5σ outlier rejection. Destriping the data also removed potential background in the 2D images, although none was apparent in the data. The jump-step and dark-current stages of the jwst pipeline[12] (v.1.6.2) were skipped, and the top and bottom six pixels of the non-illuminated subarray were manually set to be reference pixels in the jwst pipeline reference pixel step. To obtain our final wavelength calibration, we extrapolated the STScI-provided in-flight instrumental wavelength calibration data product across the detector edge pixels that did not have an assigned wavelength. The calibration was derived using the ground-based wavelength solution. We performed tests to search for zero-point offsets in the calibration versus the planetary and stellar spectra and did not find any at the level of half a pixel width or greater.

JWST detectors integrate using a non-destructive up-the-ramp sampling technique in which the flux is measured in counts per second from fitting the ramp from the groups contained within each integration. Extended Data Fig. 2 shows the regions of the spectrum affected by saturation. Within a column where a pixel was marked as saturated by the pipeline in any given group, we used only the data from the preceding groups for ramp fitting and manually set an entire column of the detector as saturated if a pixel in that column was saturated. Because a small portion of the spectrum reaches our saturation threshold in the second group, this region of the spectrum only uses one group to derive a ‘ramp’. Although we were able to recover the spectra in this wavelength range by flagging and ignoring saturated pixels at the group level, we note that the data quality is lower in the saturated region than in the rest of the spectrum given the counts per second ramp was measured from fewer than the total five groups.

We measured the positional shift of the spectral trace across the detector throughout the time series using cross correlation and used them to shift-stabilize the images with flux-conserving interpolation. This procedure reduced the amplitude of position-dependent trends in the light curves. We optimized the width of our flux extraction aperture at each wavelength pixel and extracted the spectrophotometry. For each wavelength, we tested a wide range of aperture widths and determined the width that minimized the scatter of the photometry of the first 350 data points. We bin the cleaned

spectrophotometry in wavelength to create 207 variable-width spectral channels with roughly 10^5 counts per second in each bin and widths ranging from 3.3 to 60 nm. Because we use fewer groups in the saturated detector columns, our bin widths are larger by a factor of a few in this region to account for the lower count rates per detector column.

Before fitting the transmission spectrum, we use a very wide, high-SNR white light channel (3–5.5 μm) to fit for the planet's orbital parameters (listed in Extended Data Table 1). Restricting the wide bin to the reddest wavelengths minimizes the impact of limb darkening on the transit light curve and the resulting covariance with the orbital system parameters while ignoring the saturated region. We fit this white light curve using the Markov chain Monte Carlo sampler emcee[32] within the least-squares minimization framework of lmfit. We use 1,000 steps and uniform priors with extremely wide bounds that encapsulate the limits of physicality to ensure that there is no bias introduced by the prior. Our fitting approach accounts for non-Gaussian degeneracies in the posterior distribution, thereby addressing the known linear correlation between impact parameter (b) and the scaled semimajor axis (a/R_*)

We excluded the first 3,000 integrations as they showed a slight non-linear baseline flux trend, and integrations 20,750–20,758 because of a high-gain antenna move that was identified from outliers in the photometry that correlated with noticeable trace shifts in the x and y directions. To measure the transmission spectrum, we fit the light curve at each wavelength channel jointly with a transit model[33] and a linear combination of systematics vectors composed of the measured spectral shifts in the x and y directions. At each channel, we fit the planet's transit depth and the stellar limb darkening, while fixing the transit centre time T_0, affect parameter b and normalized semimajor axis a/R_* to the values determined in the white light curve fit. We also fix the orbital period to the published value of 4.0552941 days (ref. [34]). With the orbital system parameters fixed, we find the posterior distribution is well-fit by a multivariate Gaussian distribution, and therefore use a Levenberg–Marquart least-squares minimization algorithm[24] to efficiently determine the best-fit parameters. In each channel, we inflate the transit depth error bars in quadrature with the measured residual red noise in the photometry as measured by the binning technique[35]. Measured uncertainties

on the transit depths vary from 50 to 200 ppm, with a median of 99 ppm (Extended Data Fig. 4). As the noise levels are very close to the limit with what is expected including only photon and read noise sources, tools such as PandExo[36] should accurately predict what is achievable for other planets. We measure an increase in red noise for a few select spectral channels, but otherwise the light curves show no significant systematic errors, with some channels binning down to precision levels of a few ppm. We measure x and y jitter systematics at the roughly 100 ppm level. We see differences in the central transit time as a function of wavelength on the order of 10 s, which may be attributable to limb asymmetries in the atmospheric temperature and composition. We show these signatures in Extended Data Fig. 3. Notably, we see a significant timing structure in the 2–3 µm range, which may arise from limb asymmetries in temperature and/or cloud coverage at the altitude probed by the water vapour absorption feature at 2.7 µm (ref. [37]). Further analysis of the spectrophotometry could be warranted to investigate limb asymmetries in more detail.

We fit the transit light curves using a quadratic function to model stellar limb darkening given as,

$$\frac{I\left(\mu \right)}{I\left(1\right)}=1-a\left(1-\mu \right)-b{\left(1-\mu \right)}^{2}$$
(1)

where $I(1)$ is the intensity at the centre of the stellar disc, $\mu = \cos(\theta)$ where θ is the angle between the line of sight and the emergent intensity, and a and b are the limb darkening coefficients. We tested a four-parameter non-linear limb darkening function[38] as well, which provided equivalent results. In practice, we first fit for both $u_+ = a + b$ and $u_- = a - b$ for the quadratic law. When comparing the limb darkening coefficients to theoretical values, we find an offset between the theoretically derived values of $u+$ from the 3D stellar models from ref. [39] and the JWST values derived from the transit light curve fits (Extended Data Fig. 5). This offset suggests the limb of WASP-39A is brighter than the stellar models predict. We fit for this offset and find it to be -0.065 ± 0.022. As the wavelength-to-wavelength shape of u_+ is well described by the model, we then apply this offset to the theoretical limb darkening coefficients and then subsequently fix u_+ while allowing

only u_{-} to be free (Extended Data Fig. 5). This procedure helps reduce degeneracies when fitting several limb darkening coefficients and increases the precision of the transmission spectrum, as the limb darkening is often not well constrained, particularly at long wavelengths where the limb darkening is weak[39] (Extended Data Fig. 5). The main effect of fitting for limb darkening over fixing the coefficients to the 3D models is the transit depth level of the optical spectrum, which is lower with values fixed to the model. We compare the optical spectrum with fixed limb darkening to the HST data from ref. [6] in Extended Data Fig. 6, which was also fit with limb darkening fixed to the same model. Overall, we find good agreement between the two spectra. We note that the assumptions around limb darkening can affect the optical spectra continuum wthat particularly affects the interpreted levels of aerosol scattering: further investigations are warranted.

tshirt

We use the tshirt pipeline, for example, ref. [40], to extract an independent set of light curves and spectrum. We begin with the uncalibrated 'uncal' data product and apply a custom set of processing steps on stage 1 that build on the existing jwst stage 1 pipeline software v.1.6.0 with reference files CRDS (Calibration Reference Data System) jwst_0930.pmap. We use a custom bias file shared by the instrument team (S. Birkmann, private communication), which is the same file that was delivered to the JWST CRDS.

We attempt to minimize the biasing effects of count rate non-linearity by modifying the quality flags of pixels surpassing 90% of full-well depth at the group stage. To ensure that there are no systematic differences between pixels within the spectral trace and in the background region, we adjust the quality flags uniformly along the entire pixel column at each group for all integrations. We skip the 'jump' and 'dark' steps of stage 1.

The tshirt code includes a row, odd-even by amplifier correction to reduce $1/f$ noise. We first identify source pixels by choosing pixels with more than five data numbers per second (DN s^{-1}) in the rate file and expanding this region out by 8 pixels. We then identify background pixels for $1/f$ corrections by choosing all non-source pixels and pipeline flagged non-'DO NOT USE' pixels. We loop through every group and subtract the median of odd (even)

row background pixels from all odd (even) rows. We next find a column-by-column median of all background pixels to calculate a $1/f$ stripe correction and subtract this from each column.

After calculating rate files in DN s^{-1}, we use tshirt to perform covariance-weight extraction of the spectrum[31]. We do a column-by-column linear background subtraction using pixels 0–7 and 25–32. We use a rectangular source extraction region centred on $Y = 16$ pixels with a width of 14 pixels. We assume the correlation between pixels to be 8% from previous studies of background pixels[31]. We use a spline with 30 knots to estimate a smooth spectrum of the star at the source pixels and identify bad pixels as ones that deviate by more than 50σ from the spline. Pixels that are more than 50σ or else marked as DO NOT USE are flagged and then the spatial profile is interpolated over those pixels. No corrections were made to the centroid or wavelength solution because of the exceptional pointing stability of the observatory[41].

When fitting the light curves, we exclude all time samples between UT 2022-07-10T23:20:01 and 2022-07-10T23:21:08 to avoid the effects of the high-gain antenna move. We first fit the broadband light curve with all wavelengths. We assume zero eccentricity and the orbital parameters from ref. [34] for a/R_* and period. We try fitting the white light curve with eccentricity and argument of periastron set free and find that eccentricity is consistent with 0. We therefore assume zero eccentricity and a transit centre projected to the time of observations from a fit to the TESS data. We also assume an exponential temporal baseline in time to the data and a second-order polynomial trend in time. We fit the quadratic limb darkening parameters with uninformative priors[42] and the exoplanet code[43,44,45] with 3,000 burn-in steps and 3,000 sampling steps and two No U Turns Sampling chains[46]. We next binned the spectra into 116 bins, each 4 pixels wide. We fit all the individual spectroscopic channels with the orbital parameter fixed from the broadband light curve fit and only allowed the transit depth and limb darkening parameters to be free. Our resulting transit depth uncertainties ranged from 35 to 732 ppm, with a median of 90 ppm.

Eureka! and ExoTEP

We use the Eureka! pipeline[47] for the data reduction steps of detector processing, data calibration and stellar spectrum extraction, and the ExoTEP pipeline[48,49,50] to generate light curves in each wavelength bin and perform light curve fitting.

We start our data reduction using the uncalibrated uncal outputs of the jwst pipeline's stage 0. From there, Eureka! acts as a wrapper for the first two stages of the jwst pipeline, v.1.6.0. We use the jwst pipeline to fit slopes to the ramp in each pixel and perform data calibration, and follow the default pipeline steps unless otherwise stated. We skip the jump detection step, meant to correct the ramps for discontinuities in the slopes of group count rates as a function of time. Owing to the small number of groups up the ramp, performing this step leads to a large fraction of the detector pixels being incorrectly flagged as outliers and we therefore rely on the time series outlier-clipping steps in the subsequent stages to correct for cosmic rays. A custom bias frame is used, rather than the default one available on CRDS at the time of reduction. We also expand the saturation flags in stage 1 to ignore saturated pixels more conservatively than allowed by the default jwst pipeline settings: in each group, we flag pixels as saturated if they reach roughly 85% of the full well in the median image across all integrations for that group and expand the saturation flag such that in a given detector column (constant wavelength) all pixels are marked as saturated if any one pixel in that column is flagged. This is implemented by inputting the indices of columns to mask on the basis of an inspection of the uncal data products, rather than an internal calculation of the full well percentage. We include a version of the row, odd-even by amplifier correction described above, using the top and bottom six rows. We further add a custom background correction at the group level before ramp fitting, and subtract from each column the median of the six pixels at the top and at the bottom of the detector, excluding outliers at more than the 3-σ level. We skip the 'photom' step in stage 2 of the STScI detector pipeline because absolute fluxes are not needed in our analysis. We also skip the 'extract1d' step as we perform custom spectral extraction using Eureka!.

For 1D spectral extraction, we trim the array to include only columns 14 to 495 in the dispersion direction, as NIRSpec's throughput is negligible beyond this range. We then use the median detector frame to construct the

weights used in the optimal extraction based on ref. [51]. Pixels are masked if they have an marked data quality flag (that is, bad pixels that are flagged by the jwst pipeline as 'DO NOT USE' for various reasons) or if they are clipped by two iterations of 10-σ-clipping of the time series. We perform the optimal extraction over eight rows centred on the source position (corresponding to a spectral half-width aperture of 4 pixels). The source position is identified from the maximum of a Gaussian fitted to the summed spatial profile from all detector columns over the entire integration.

We use ExoTEP to generate median-normalized light curves at the native pixel resolution from each detector column, using the stellar spectra outputs from stage 3 of Eureka!. We then perform further clipping of outliers in time in the white and wavelength-dependent light curves by computing a running median with a window size of 20 and excluding 3σ outliers in several time series. This outlier-clipping was applied to the flux, source position and width in the cross-dispersion direction in each frame and spectrum shifts in the dispersion direction.

We jointly fit astrophysical and systematics model parameters to the white (0.5–5.5 µm) light curves and each of the wavelength-dependent light curves. Our astrophysical transit model is calculated using the batman package[33]. Using the white light curve, we fit for the two coefficients of a quadratic limb darkening law (equation (1)), WASP-39b's impact parameter, scaled semimajor axis a/R_*, time of transit centre and the planet-to-star radius ratio. In each of the wavelength channels we then fix the planet's impact parameter, semimajor axis and transit time to the values derived from the white light curve and fit only for the planet-to-star radius ratio and the two quadratic limb darkening coefficients. For the systematics model, we assume a linear trend with time that can be different in each spectroscopic channel, and fit for its slope and y intercept. Last, we fit a single-point scatter to each light curve, which illustrates the level of scatter required for our joint model to reach a reduced chi-squared of 1. The fitted light curve scatter in both the white light curve and wavelength-dependent channels is within a few percent of the expectation from the high-frequency scatter in the raw light curves, which attests to the lack of systematics. We bin the final transmission spectrum (four points binned together throughout the spectrum) for visual comparison with the other reductions in Fig. 3.

Tiberius

The Tiberius pipeline builds on the LRG-BEASTS spectral reduction and analysis pipelines introduced in refs. [16,52,53]. The Tiberius pipeline operates on the stage 1 JWST data products to obtain 1D stellar spectra through tracing of the stellar spectra, fitting and removal of the background noise and simple aperture photometry. We used the FIREFLy-processed stage 0 data.

Before tracing the spectra, we interpolate each column of the detector onto a finer grid, 10× the initial spatial resolution. This step improves the extraction of flux at the subpixel level, particularly where the edges of the photometric aperture bisect a pixel, and leads to a 14% reduction in the noise in the data. We also interpolate over the bad pixels using their nearest neighbouring pixels in x and y. We identify bad pixels by combining 5σ outlying pixels found by means ofrunning medians operating along the pixel rows with bad pixels identified by visual inspection. We trace the spectrum by fitting a Gaussian distribution at each column (in which a column refers to the cross-dispersion direction) to the stellar spectra. We then use a running median, calculated with a moving box with a width of five data points, to smooth the measured centres of the trace. We fit these smoothed centres with a fourth-order polynomial, removed five median absolute deviation outliers and refitted with a fourth-order polynomial.

To remove residual background flux not captured by the $1/f$ correction, we fit a linear polynomial along each column in the spatial direction. We mask the stellar spectrum, defined by an aperture with a full width of 4 pixels centred on the trace we found in the previous step, from this background fit. We also mask an extra 7 pixels on either side of this aperture so that the background fit is not affected by the wings of the stellar point spread function. This left us with 7 pixels at each edge of the detector (a total of 14 pixels) with which to estimate the background. We also clipped any pixels within the background that deviate by more than three standard deviations from the mean for that particular column and frame to avoid residual bad pixels and cosmic rays affecting our background estimation. We found that this extra background step led to a 3% improvement in the precision of the transmission spectrum.

The stellar spectra are then extracted by summing the flux within a 4-pixel-wide aperture following the removal of the background at each column. The background count level, as estimated by the JWST Exposure Time Calculator is on the order of a few counts per second, meaning the background is negligible. Further, because we perform $1/f$ subtraction, this faint background is subtracted column-by-column. We experimented with the choice of the aperture width, also running reductions with 8- and 16-pixel-wide apertures. The 8-pixel-wide aperture gave a median uncertainty 1% larger than a 4-pixel aperture and a 16-pixel aperture gave an uncertainty 15% larger than 4 pixels. This same change was reflected in the median root mean square of the residuals to the light curve fits. As the stellar point spread function is so narrow in PRISM data, we believe that the increase in noise with increasing aperture width is related to the increasing influence of photon noise, read noise and bad pixels where the stellar flux is lower. Following the extraction of the stellar spectra, we divide the measured count rates by a factor of ten to correct for our pixel oversampling, as described above.

To remove residual cosmic rays, we identify outliers in each stellar spectrum through comparison with the median stellar spectrum. We did this in three iterations, each of which involves making a median spectrum, identifying outliers (10, 9, 8 σ) and replacing pixels containing a cosmic ray with a linear interpolation between neighbouring pixels. We tested this interpolation against assigning the cosmic ray pixels zero weight and found that this led to a negligible difference in the transmission spectrum. To correct for shifts in the stellar spectra and align each spectrum in pixel space, we cross-correlate each stellar spectrum with the first spectrum of the observation and linearly resample each spectrum onto a common wavelength grid. We adopt the custom wavelength solution calculated by the tshirt pipeline, which uses the jwst pipeline to evaluate the wavelengths at pixel row 16 using the world coordinate system.

Our white light curves are created by summing over the full wavelength range between 0.518 and 5.348 µm. We make two sets of spectroscopic light curves: one set of 440 light curves at 1-pixel resolution and one set of 147 light curves at 3-pixel resolution. We mask integrations 20,751–20,765 because of a high-gain antenna move that leads to increased noise in the light curves. We also mask the first 2,000 integrations from our analysis

because of a systematic ramp. This means our light curves each contained 19,486 data points.

To fit our light curves, we began by fitting the white light curve to determine the system parameters.

We fit for the following parameters: the scaled planetary radius (R_p/R_*), the planet's orbital inclination (i), the time of mid-transit (T_C), the scaled separation (a/R_*), the linear limb darkening coefficient (u_1) and the parameters defining the systematics model. We fix the planet's orbital period to 4.0552941d and eccentricity to 0 (ref. [34]). For the remaining parameters, we use the values from ref. [34] as initial guesses.

For the analytic transit light curve model, we use batman[33] with a quadratic limb darkening law. We use ExoTiC-LD[54,55], with 3D stellar models[39] to determine the appropriate limb darkening coefficients (LDs), adopting the stellar parameters (T_{eff} = 5,512 ± 55 K, log g = 4.47 ± 0.03 cgs, [Fe/H] = 0.01 ± 0.09 dex) from ref. [34] and Gaia DR3 (refs. [56,57]). For our final fits, we fix the quadratic coefficient, u_2, to the values determined by ExoTiC-LD. However, we also run a set of fits with neither u_1 nor u_2 fixed and find this leads to a transmission spectrum that is qualitatively similar to the one in which LDs are fixed. For the systematics model, we sum the following three polynomials: quadratic in time, linear in x position of the star on the detector and linear in y position of the star on the detector. The final fit model, M, was of the form:

$$M(t)=T(t,p)\times ({\sum }_{i}({S}_{i}{({a}_{i},s)}^{ni}))$$
(2)

where t is time, p are the parameters of the transit model, T, a are the ancillary data and s are the parameters (polynomial coefficients) of the systematics model, S. The systematics model is the sum of the polynomials operating over each ancillary input, a_i, with n_i defining the order of the polynomial used for each input.

We fit our white light curve in three steps: a first fit to remove any 4σ outliers from the light curves, a second fit that is used to rescale the

photometric uncertainties such that the best-fitting model gives $\chi_v^2 = 1$ and a third fit with the rescaled photometric uncertainties, from which our final parameter values and uncertainties are estimated. The parameter uncertainties were calculated as the standard deviation of the diagonal of the covariance matrix that was in turn calculated from the Jacobian returned by scipy.optimize.

Following the white light curve, we fit our spectroscopic, wavelength-binned, light curves. For these fits, we held a/R_*, i and T_C fixed to the values determined from the white light curve fit: 11.462 ± 0.014, 87.847 ± 0.015°, 2,459,770.835623 ± 0.000008 Barycentric Julian Date Dynamical Time (BJD_{TDB}.). These values are different from the FIREFLy-reduced white light parameters, and these differences will be explored in greater detail in a future work. To zeroth order, offsets in orbital parameters result in simple vertical offsets in the resulting transmission spectrum. The remaining fit parameters were the same as for the white light curve fit. We perform the same iteration of fits using a Levenberg–Marquardt algorithm to determine R_p/R_s as a function of wavelength.

Reduction comparison

Procedural differences exist across the four main reductions of the dataset, which may account for the subtle qualitative differences between the final reduced spectra. A careful investigation of these nuances is warranted and will be presented in a future paper. Extended Data Table 2 highlights some key procedural differences between the reductions. We note that, despite these differences, the resulting exoplanet spectra are qualitatively in excellent agreement with each other (Fig. 3) because of the stability of the data and the self-calibrating nature of the transit technique.

Stellar activity

WASP-39b has a reported low activity level[8], with a Ca II H and K stellar activity index of $\log R'_{HK} = -4.994$ (ref. [4]). NGTS and TESS photometric monitoring of WASP-39A is reported in ref. [22], which finds low modulations at the 0.06% level with no apparent star-spot crossings. With low stellar

activity levels, the transit observations are unlikely to be affected by stellar activity.

Forward model grids

We use four different 1D RCTE model grids to assess atmospheric properties such as detection of individual gases, metallicity, carbon-to-oxygen (C/O) elemental abundance ratio, and the presence/absence of clouds. The ScCHIMERA[58,59], PICASO 3.0 (refs. [60,61,62,63]), ATMO[54,64,65] and PHOENIX[66,67] models were used to generate these grids specifically for WASP-39b. Whereas the ATMO and the PHOENIX grids were used to fit the data with a reduced χ^2 based grid search method, the PICASO 3.0 and ScCHIMERA grids were used in a grid retrieval framework using a nested sampler[68,69]. Within each nested sample likelihood calculation, the transmission spectra are generated on-the-fly by postprocessing the precomputed 1D RCTE model atmospheres. The SO_2 volume mixing ratio and cloud properties are injected into spectrum during this postprocessed transmission calculation. Extended Data Fig. 7 shows best-fit models obtained by each of the four grids compared with the transmission spectrum obtained with the FIREFLy data reduction pipeline. ScCHIMERA, PICASO 3.0 and ATMO produce fits with reduced χ^2 between 3.2 and 3.3, while the PHOENIX grid obtains a reduced χ^2 of 4.3. The reduced χ^2 is defined as the total χ^2 calculated from all the data points divided by the total number of data points. Although PICASO 3.0, ScCHIMERA and ATMO predict the metallicity of the atmosphere to be about 10× solar, PHOENIX finds a best-fit metallicity to be a 100× solar that might be due to the larger grid spacing of the PHOENIX grid along both the cloud and metallicity dimensions. Although the models qualitatively match the data, the reduced χ^2 obtained by the best-fitting models from these grids are also greater than three, which suggests that these are not fitting the data particularly well. These poor fits could arise for many reasons, such as the region of the data affected by saturation, the presence of disequilibrium chemistry in the atmosphere due to vertical mixing or photochemistry and the non-grey nature of scattering in the upper atmosphere. Extended Data Table 3 provides a summary of the best-fit atmospheric parameters obtained by the four different grids with different fitting methods (grid retrievals and grid search). To explore the

effect of the saturated region on the best-fit parameters, we inflate the transit depth errors in the saturated regions (0.68–1.91 μm) by a factor of 1,000 and recompute the best-fit models using the grid retrieval framework with both the PICASO v.3.0 and ScCHIMERA grids. We find that this did not significantly change any of the best-fit parameters including the metallicity and the C/O ratio. Extended Data Table 3 lists the best-fit parameters obtained when the saturated region error bars were inflated by a factor of 1,000. We summarize the main results obtained by these 1D grids here and refer the reader to ref. [29] for detailed descriptions of each of these model grids.

Detection significance of gases

We quantify the detection significance of each species through a Bayes factor analysis, for example ref. [70]. To do so within the ScCHIMERA grid retrieval framework, we remove each gas during the transmission spectrum computation step (the 1D RCTE atmosphere models remain unchanged) one at a time and re-run the nested sampler. We compare the Bayesian evidence of each removed-gas run to that of grid retrieval with all the gases. There is no change in the number of parameters except the cloud and SO_2 mixing ratio parameters. Extended Data Table 4 shows the result of this exercise summarized as the log-Bayes factor and a conversion to the detection significance: for example ref. [71].

We also quantify the detection significances of different gases following the procedure used in ref. [29]. To calculate the detection significance of each gas, the best-fit transmission spectrum model from the PICASO 3.0 grid ([M/H] = +1.0, C/O = 0.68) is recalculated without that gas. The wavelength ranges in which the particular gas has the most prominent effect are first identified and then a residual spectrum is calculated by subtracting the model without the gas from the data. The residual spectra for H_2O, CO_2, CO, Na, SO_2 and CH_4 are shown in the six panels of Extended Data Fig. 8. We fit each of these residual spectra with two functions, a Gaussian–double Gaussian–Voigt function and a constant line. We use the Dynesty nested sampling routine[68] to perform the fits and to determine the Bayesian evidence associated with each fit. The Bayes factor between the fits of the residual

spectrum with the Gaussian–Voigt function and the constant line is then used to determine the detection significance of a gas. For example, for computing the detection significance of H_2O, two adjacent H_2O features between 1 and 2.2 µm are used. We note that H_2O is expected to be the dominant opacity source in other wavelength ranges (for example, 2.2–3 µm) as well, so choosing two features for this analysis would produce a lower limit on the detection significance of H_2O. The best-fit double Gaussian function to these features along with its 1σ and 2σ envelopes are shown with the red line and shaded regions in Extended Data Fig. 8 top-left panel. The same residual spectrum is also fitted with a straight line shown with blue colour in Extended Data Fig. 8. The logarithm of the Bayes factor between the two models is found to be $\ln B = 242$, which shows that the model with H_2O is significantly favoured over a model without any H_2O. The detection significance of H_2O corresponding to this Bayes factor is calculated using the prescription in ref. [71] and is found to be 22σ. The same methodology, but with a single Gaussian function, is also followed for CO_2, CO, SO_2, H_2S and CH_4 to get their detection significance summarized in Extended Data Table 4, in the last column. Our Gaussian residual fit significance for CO_2 matches the initial analysis of the NIRSpec PRISM data presented in ref. [29].

As shown in Extended Data Table 4, the detection significance of all gases increases with the Bayes factor analysis technique relative to the Gaussian–Voigt function technique. This is notably also the case for SO_2, lending confidence to the detection and identification of the molecule, as the feature is better fit by its respective opacity profile.

Resolution bias and the detection significance of CO

The resolution-linked bias effect serves to dilute the measured amplitudes of planetary atmospheric features because of overlapping absorption lines in the stellar atmosphere. Although this effect is negligible for most stars earlier than M dwarfs, some stellar CO absorption is expected in WASP-39, meaning the measured planetary CO abundance may be biased. Following equation 4 in ref. [72] and using high-resolution (R of roughly 10^5) PHOENIX models of the planet and the star, we quantify an upper limit on the

magnitude of this bias effect. We find that the planetary CO feature is biased by 30 to 40 ppm in the 4.5–5.1 μm region, leading to as much as a roughly 1 $-\sigma$ underestimate of the planetary CO absorption strength, and subsequently a similar underestimate of its abundance. We note that this effect is potentially weakened by Doppler broadening of the molecular lines (which is unaccounted for by PHOENIX) because of stellar rotation, planetary orbital radial velocity and planetary winds. Future work, which may benefit from more detailed modelling and high-resolution observations of WASP-39's CO band heads, will better quantify the magnitude of this dilution.

Metallicity, C/O ratio and CH_4 abundance

The best-fitting atmospheric metallicity for WASP-39b is found to be roughly ten times the solar metallicity using the model grids. The top panel in Extended Data Fig. 9 shows the observed transmission spectrum of the planet between 2.0 and 5.3 μm (in which variations due to metallicity are most prominent), along with several transmission spectrum models assuming different atmospheric metallicities ranging from subsolar values (for example, 0.3× solar) to super-solar values (for example, 100× solar). The bottom panel demonstrates the effect of different atmospheric C/O ratios at ten times solar metallicity on many transmission spectrum models along with the data. As the star WASP-39 has near-solar elemental abundances[73], scaled solar abundances are a reasonable choice for this star. The CH_4 feature between 3.1–4 and 2.2–2.5 μm is very prominent in subsolar and solar metallicity thermochemical equilibrium models shown in Extended Data Fig. 9. The absence of such a CH_4 feature in the data is evident. This, combined with the large CO_2 feature between 4.3 and 4.6 μm and measurable CO feature at 4.7 μm, led to a super-solar (10×) metallicity estimate for the planet. The C/O ratio of the RCTE models significantly affects the predicted gas abundances, and therefore the calculated transmission spectrum. Extended Data Fig. 9 bottom panel shows that for metal-rich atmospheres (for example, >10× solar) with C/O ratios lower than 0.7, the transmission spectrum is dominated by features of oxygen-bearing gases (H_2O, CO_2, CO): for example, refs. [65,74,75]. But for higher C/O ratios (for example, 0.916), the transmission spectrum becomes CH_4 dominated at wavelengths greater than 1.5 μm. We obtain an upper limit on

the C/O ratio of WASP-39b at about 0.7. However, these interpretations are based on single-best fits from model grids assuming thermochemical equilibrium. Other chemical disequilibrium processes such as atmospheric mixing and high-energy stellar radiation-induced photochemistry can also potentially affect this interpretation. These disequilibrium chemistry effects require further exploration in the context of WASP-39b and will be discussed in future work (Welbanks et al. (in prep), Tsai et al. (submitted)).

The best-fitting metallicity models can be used to place an upper limit on the CH_4 abundance, if the pressure ranges probed by the transmission spectrum are estimated. To estimate the pressure ranges probed by the data, we use the best-fit PICASO 3.0 model to calculate a pressure- and wavelength-dependent transmission contribution function of the atmosphere[76]. This contribution function for the best-fit 10× solar metallicity PICASO v.3.0 model is shown as a heat map in Extended Data Fig. 10. This shows that the data mostly probes pressure ranges between 0.1 and 2 mbar. We also computed contribution functions for models with solar metallicity and find that they probe similar pressure ranges as well. Extended Data Fig. 10 also shows the pressure dependent CH_4 abundances in models with different metallicities presented in Extended Data Fig. 9 top panel. As only super-solar metallicity thermochemical equilibrium models are preferred by the data, the abundance profiles in Extended Data Fig. 10 help us in putting an upper limit of 5×10^{-6} on the CH_4 volume mixing ratio between 0.1 and 2 mbar.

Clouds

The observed spectrum shows slightly muted transit depths, across the entire wavelength range, compared with the depths expected from clear atmospheric models. This hints towards some extra opacity source in the atmosphere with weak wavelength dependence. Opacity sources such as clouds can mute the spectral features in a transmission spectrum[2,4]. We postprocess the transmission spectrum models with grey (that is, wavelength-independent) cloud opacities to check whether they are preferred over clear atmospheric models by the data. However, the treatment of clouds differ between the four 1D RCTE model grids. PICASO 3.0 and

ScCHIMERA grids implemented the cloud opacities using the following equation,

$${\tau }_{i,{\rm{cld}}}={\kappa }_{{\rm{cld}}}\frac{\delta {P}_{i}}{g}$$
(3)

where $\tau_{i,\mathrm{cld}}$ is the cloud optical depth of the *i*th atmospheric layer in the model with pressure width δP_i and g represents the gravity of the planet. The best-fit value of the grey cloud opacity $\kappa_{\mathrm{cld}} = 10^{-2.07}$ cm^2 g^{-1} is calculated in a Bayesian framework by postprocessing the RCTE model grid with this cloud opacity and comparing these postprocessed models with the data. The ATMO grid includes grey cloud decks at several pressures between 1 and 50 mbar, but with variable factors 0, 0.5, 1 and 5 governing cloud opacity with respect to H_2's scattering cross-section at 0.35 µm, where a factor 0 indicates a cloud-free model spectrum. The PHOENIX grid includes similar cloud decks but between 0.3 and 10 mbar with cloud optical depth enhancement factors (identically defined as the ATMO grid) 0 and 10. We find that the cloudy models better fit the data than clear models across all four model grids. The contribution of clouds in limiting the depths of the gaseous features across the entire wavelength range is also shown in Fig. 4 with the grey shaded region.

4 µm SO_2 feature identification

None of the 1D RCTE models capture the 4 µm absorption feature seen in the data. We searched for several candidate gas species that could produce this feature if their abundances differ from the expected abundances from thermochemical equilibrium. The list of searched chemical species include C-bearing gases such as C_2H_2, CS, CS_2, C_2H_6, C_2H_4, CH_3, CH, C_2, CH_3Cl, CH_3F, CN and CP. Various metal hydrides, bromides, flourides and chlorides such as LiH, AlH, FeH, CrH, BeH, TiH, CaH, HBr, LiCl, HCl, HF, AlCl, NaF and AlF were also searched as potential candidates to explain the feature. SO_2, SO_3, SO and SH are among the sulfur-based gases that were considered. Other species that were considered include gases such as PH_3, H_2S, HCN, N_2O, GeH_4, SiH_4, SiO, AsH_3, H_2CO, $H^+{}_3$, OH^+, KOH, Brα-H,

AlO, CN, CP, CaF, H_2O_2, H_3O^+, HNO_3, KF, MgO, PN, PO, PS, SiH, SiO2, SiS, TiO and VO.

Among all these gases, SO_2 was the most promising candidate in terms of its spectral shape and chemical plausibility, although the expected chemical equilibrium abundance of SO_2 is too low to produce the absorption signal seen in the data. However, previous work exploring photochemistry in exoplanetary atmospheres[25,26] have shown that higher amounts of SO_2 can be created in the upper atmospheres of irradiated planets through photochemical processes. Therefore, we postprocess the PICASO 3.0 and ScCHIMERA chemical equilibrium models with varying amounts of SO_2 in a Bayesian framework to estimate the SO_2 abundance required to explain the strength of the 4-µm feature. The required volume mixing ratio of SO_2 was found to be roughly 10^{-5}–10^{-6}. Note that in obtaining this estimate we assumed that the SO_2 volume mixing ratio does not vary with pressure for simplicity. In a photochemical scenario this assumption is probably not realistic, although the pressure range probed by SO_2 is also limited. Whether photochemical models can produce this amount of SO_2 in the atmospheric conditions of WASP-39b is a pressing question that the ERS team is now exploring (Welbanks et al. (in prep), Tsai et al. (submitted)). Whether this feature can be better explained by any other gaseous absorber is also at present under investigation by the ERS team.

Data availability

The data used in this paper are associated with JWST program ERS 1366 and are available from the Mikulski Archive for Space Telescopes (https://mast.stsci.edu). The data products required to generate Figs. 1, 2, 3, and Extended Data Figs. 1, 3 and 5 are available here: https://zenodo.org/record/7388032. All additional data are available upon request.

Code availability

The codes used in this publication to extract, reduce, and analyse the data are as follows; STScI JWST Calibration pipeline (https://github.com/spacetelescope/jwst), FIREFLy[23], tshirt[40], Eureka![47] (https://eurekadocs.readthedocs.io/en/latest/) and Tiberius[16,52,53]. In addition, these made use of Exoplanet[43] (https://docs.exoplanet.codes/en/latest/), Pymc3 (ref. [77]) (https://docs.pymc.io/en/v3/index.html), ExoTEP[48,49,50], Batman[33], (http://lkreidberg.github.io/batman/docs/html/index.html), ExoTiC-ISM[54] (https://github.com/Exo-TiC/ExoTiC-ISM), ExoTiC-LD[55] (https://exotic-ld.readthedocs.io/en/latest/), Emcee[32] (https://emcee.readthedocs.io/en/stable/), Dynesty[68] (https://dynesty.readthedocs.io/en/stable/index.html) and chromatic (https://zkbt.github.io/chromatic/), which use the python libraries scipy[78], numpy[79], astropy[80,81] and matplotlib[82]. The atmospheric models used to fit the data can be found at PICASO[60,61,62,63] (https://natashabatalha.github.io/picaso/), Virga85 (https://natashabatalha.github.io/virga/), ScCHIMERA[58,59] (https://github.com/mrline/CHIMERA), ATMO[65,83] and PHOENIX[66,67].

References

1. Seager, S. & Sasselov, D. D. Theoretical transmission spectra during extrasolar giant planet transits. *Astrophys. J.* **537**, 916–921 (2000).

2. Brown, T. M. Transmission spectra as diagnostics of extrasolar giant planet atmospheres. *Astrophys. J.* **553**, 1006–1026 (2001).

3. Hubbard, W. B. et al. Theory of extrasolar giant planet transits. *Astrophys. J.* **560**, 413–419 (2001).

4. Sing, D. K. et al. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature* **529**, 59–62 (2016).

5. Zhang, X. Atmospheric regimes and trends on exoplanets and brown dwarfs. *Res. Astron. Astrophys.* **20**, 099 (2020).

6. Wakeford, H. R. et al. The complete transmission spectrum of WASP-39b with a precise water constraint. *Astron. J* **155**, 29 (2018).

7. Mikal-Evans, T. et al. Transmission spectroscopy for the warm sub-Neptune HD 3167c: evidence for molecular absorption and a possible high-metallicity atmosphere. *Astron. J* **161**, 18 (2021).

8. Faedi, F. et al. WASP-39b: a highly inflated Saturn-mass planet orbiting a late G-type star. *Astron. Astrophys.* **531**, A40 (2011).

9. Birkmann, S. M. et al. The near-infrared spectrograph (NIRSpec) on the James Webb Space Telescope IV. Capabilities and predicted performance for exoplanet characterization. *Astron. Astrophys.* **661**, A83 (2022).

10. Stevenson, K. B. et al. Transiting exoplanet studies and community targets for JWST's Early Release Science program. *Publ. Astron. Soc. Pacif.* **128**, 094401 (2016).

11. Bean, J. L. et al. The transiting exoplanet community Early Release Science program for JWST. *Publ. Astron. Soc. Pac.* **130**, 114402 (2018).

12. Bushouse, H. et al. JWST Calibration Pipeline. *Zenodo* https://doi.org/10.5281/zenodo.6984366 (2022).

13. Ferruit, P. et al. The near-infrared spectrograph (NIRSpec) on the James Webb Space Telescope II. Multi-object spectroscopy (MOS). *Astron. Astrophys.* **661**, A81 (2022).

14. Fischer, P. D. et al. HST hot-Jupiter transmission spectral survey: clear skies for cool Saturn WASP-39b. *Astrophys. J.* **827**, 19 (2016).

15. Nikolov, N. et al. VLT FORS2 comparative transmission spectroscopy: detection of Na in the atmosphere of WASP-39b from the ground. *Astrophys. J.* **832**, 191 (2016).

16. Kirk, J. et al. LRG-BEASTS: transmission spectroscopy and retrieval analysis of the highly inflated Saturn-mass planet WASP-39b. *Astron. J*

158, 144 (2019).

17. Tsiaras, A. et al. A population study of gaseous exoplanets. *Astron. J* **155**, 156 (2018).

18. Fisher, C. & Heng, K. Retrieval analysis of 38 WFC3 transmission spectra and resolution of the normalization degeneracy. *Mon. Not. R. Astron. Soc.* **481**, 4698–4727 (2018).

19. Pinhas, A. et al. H_2O abundances and cloud properties in ten hot giant exoplanets. *Mon. Not. R. Astron. Soc.* **482**, 1485–1498 (2019).

20. Welbanks, L. et al. Mass-metallicity trends in transiting exoplanets from atmospheric abundances of H_2O, Na, and K. *Astrophys. J. L.* **887**, L20 (2019).

21. Thorngren, D. P. et al. The mass-metallicity relation for giant planets. *Astrophys. J.* **831**, 64 (2016).

22. Ahrer, E.-M. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRCam. *Nature* https://doi.org/10.1038/s41586-022-05590-4 (2023).

23. Rustamkulov, Z. A. et al. Analysis of a JWST NIRSpec lab time series: characterizing systematics, recovering exoplanet transit spectroscopy, and constraining a noise floor. *Astrophys. J. L.* **928**, L7 (2022).

24. Newville, A. et al. LMFIT: non-linear least-square minimization and curve-fitting for Python. *Zenodo* https://doi.org/10.5281/zenodo.11813 (2014).

25. Tsai, S.-M. et al. A comparative study of atmospheric chemistry with VULCAN. *Astrophys. J.* **923**, 264 (2021).

26. Zahnle, K. et al. Atmospheric sulfur photochemistry on hot Jupiters. *Astrophys. J. L.* **701**, L20–L24 (2009).

27. Hobbs, R. et al. Sulfur chemistry in the atmospheres of warm and hot Jupiters. *Mon. Not. R. Astron. Soc.* **506**, 3186–3204 (2021).

28. Alderson, L. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H. *Nature* https://doi.org/10.1038/s41586-022-05591-3 (2023).

29. JWST Transiting Exoplanet Community Early Release Science Team. Identification of carbon dioxide in an exoplanet atmosphere. *Nature* https://doi.org/10.1038/s41586-022-05269-w (2022).

30. Dobbs-Dixon, I., Agol, E. & Burrows, A. The impact of circumplantary jets on transit spectra and timing offsets for hot Jupiters. *Astrophys. J.* **751**, 87 (2012).

31. Schlawin, E. et al. JWST noise floor. I. Random error sources in JWST NIRCam time series. *Astron. J* **160**, 231 (2020).

32. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306 (2013).

33. Kreidberg, L. batman: basic transit model calculation in Python. *Publ. Astron. Soc. Pacif.* **127**, 1161 (2015).

34. Mancini, L. et al. The GAPS programme with HARPS-N at TNG. XVI. Measurement of the Rossiter McLaughlin effect of transiting planetary systems HAT-P-3, HAT-P-12, HAT-P-22, WASP-39, and WASP-60. *Astron. Astrophys.* **613**, A41 (2018).

35. Pont, F., Zucker, S. & Queloz, D. The effect of red noise on planetary transit detection. *Mon. Not. R. Astron. Soc.* **373**, 231–242 (2006).

36. Batalha, N. E. et al. PandExo: a community tool for transiting exoplanet science with JWST & HST. *Publ. Astron. Soc. Pacif.* **129**, 064501 (2017).

37. Powell, D. et al. Transit signatures of inhomogeneous clouds on hot Jupiters: insights from microphysical cloud modeling. *Astrophys. J.* **887**, 170 (2019).

38. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for $-5.0 <= \log[M/H] <= +1$, 2000 K

<= Teff <= 50000 K at several surface gravities. *Astron. Astrophys.* **363**, 1081–1190 (2000).

39. Magic, Z., Chiavassa, A., Collet, R. & Asplund, M. The Stagger-grid: a grid of 3D stellar atmosphere models. IV. Limb darkening coefficients. *Astron. Astrophys.* **573**, A90 (2015).

40. Glidic, K. et al. Atmospheric characterization of hot Jupiter CoRoT-1 b using the wide field camera 3 on the Hubble Space Telescope. *Astron. J* **164**, 19 (2022).

41. Rigby, J. et al. Characterization of JWST science performance from commissioning. Preprint at https://arxiv.org/abs/2207.05632 (2022).

42. Kipping, D. M. Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. *Mon. Not. R. Astron. Soc.* **435**, 2152–2160 (2013).

43. Foreman-Mackey, D. et al. exoplanet: Gradient-based probabilistic inference for exoplanet data & other astronomical time series. *J. Open Source Softw.* **6**, 3285 (2021).

44. Luger, R. et al. starry: analytic occultation light curves. *Astron. J* **157**, 64 (2019).

45. Agol, E., Luger, R. & Foreman-Mackey, D. Analytic planetary transit light curves and derivatives for stars with polynomial limb darkening. *Astron. J* **159**, 123 (2020).

46. Hoffman, M. D. & Gelman, A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Preprint at https://arxiv.org/abs/1111.4246 (2011).

47. Bell, T. J. et al. Eureka!: An end-to-end pipeline for JWST time-series observations. *J. Open Source Softw.* **7**, 4503 (2022).

48. Benneke, B. et al. Spitzer observations confirm and rescue the habitable-zone super-Earth K2-18b for future characterization. *Astrophys. J.* **834**, 187 (2017).

49. Benneke, B. et al. A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds. *Nat. Astron.* **3**, 813–821 (2019).

50. Benneke, B. et al. Water vapor and clouds on the habitable-zone sub-Neptune exoplanet K2-18b. *Astrophys. J. L.* **887**, L14 (2019).

51. Horne, K. An optimal extraction algorithm for CCD spectroscopy. *Publ. Astron. Soc. Pacif.* **98**, 609–617 (1986).

52. Kirk, J. et al. Rayleigh scattering in the transmission spectrum of HAT-P-18b. *Mon. Not. R. Astron. Soc.* **468**, 3907–3916 (2017).

53. Kirk, J. et al. ACCESS and LRG-BEASTS: a precise new optical transmission spectrum of the ultrahot Jupiter WASP-103b. *Astron. J* **162**, 34 (2021).

54. Laginja, I. & Wakeford, H. R. ExoTiC-ISM: a Python package for marginalisExoplanet transit parameters across a grid of systematic instrument models. *J. Open Source Softw.* **5**, 2281 (2020).

55. Wakeford, H. & Grant, D. Exo-TiC/ExoTiC-LD:ExoTiC-LD v.2.1 Zenodo Release. *Zenodo* https://zenodo.org/record/6809899 (2022).

56. Gaia Collaboration. et al. The Gaia mission. *Astron. Astrophys.* **595**, A1 (2016).

57. Gaia Collaboration. et al. Gaia Early Data Release 3. Summary of the contents and survey properties. *Astron. Astrophys.* **649**, A1 (2021).

58. Piskorz, D. et al. Ground- and space-based detection of the thermal emission spectrum of the transiting hot Jupiter KELT-2Ab. *Astron. J* **156**, 133 (2018).

59. Arcangeli, J. et al. H^- opacity and water dissociation in the dayside atmosphere of the very hot gas giant WASP-18b. *Astrophys. J. L.* **855**, L30 (2018).

60. Mukherjee, S., Batalha, N. E., Fortney, J. J. & Marley, M. S. PICASO 3.0: a one-dimensional climate model for giant planets and brown dwarfs. *Astrophys. J.* **942**, 2 (2022).

61. Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D. & Freedman, R. Comparative planetary atmospheres: models of TrES-1 and HD 209458b. *Astrophys. J. L.* **627**, L69–L72 (2005).

62. Marley, M. S. et al. The Sonora brown dwarf atmosphere and evolution models. I. Model description and application to cloudless atmospheres in Rainout chemical equilibrium. *Astrophys. J.* **920**, 85 (2021).

63. Marley, M. S. & McKay, C. P. Thermal structure of Uranus' atmosphere. *Icarus* **138**, 268–286 (1999).

64. Tremblin, P. et al. Fingering convection and cloudless models for cool brown dwarf atmospheres. *Astrophys. J. L.* **804**, L17 (2015).

65. Goyal, J. M. et al. A library of self-consistent simulatExoplanet atmospheres. *Mon. Not. R. Astron. Soc.* **498**, 4680–4704 (2020).

66. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885–895 (2001).

67. Lothringer, J. D. & Barman, T. S. The PHOENIX exoplanet retrieval algorithm and using H^- Opacity as a probe in ultrahot Jupiters. *Astron. J* **159**, 289 (2020).

68. Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158 (2020).

69. Buchner, J. et al. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astron. Astrophys.* **564**, A125 (2014).

70. Benneke, B. & Seager, S. How to distinguish between cloudy mini-Neptunes and water/volatile dominated super-Earths. *Astrophys. J.* **778**, 153 (2013).

71. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104 (2008).

72. Deming, L. D. & Seager, S. Illusion and reality in the atmospheres of exoplanets. *J. Geophys. Res. Planets* **122**, 53–75 (2017).

73. Polanski, A. S., Crossfield, I. J. M., Howard, A. W., Isaacson, H. & Rice, M. Chemical abundances for 25 JWST exoplanet host stars with KeckSpec. *Res. Notes AAS* **6**, 155 (2022).

74. Madhusudhan, N. C/O ratio as a dimension for characterizing exoplanetary atmospheres. *Astrophys. J.* **758**, 36 (2012).

75. Drummond, B. et al. The carbon-to-oxygen ratio: implications for the spectra of hydrogen-dominatExoplanet atmospheres. *Mon. Not. R. Astron. Soc.* **486**, 1123–1137 (2019).

76. Molli`ere, P. et al. petitRADTRANS. A Python radiative transfer package for exoplanet characterization and retrieval. *Astron. Astrophys.* **627**, A67 (2019).

77. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using pymc3. *PeerJ Comp. Sci.* **2**, e55 (2016).

78. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

79. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

80. Astropy Collaboration. et al. Astropy: a community Python package for astronomy. *Astron. Astrophys.* **558**, A33 (2013).

81. Astropy Collaboration. et al. The Astropy project: building an open-science project and status of the v2.0 core package. *Astron. J* **156**, 123 (2018).

82. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

83. Goyal, J. M. et al. A library of ATMO forward model transmission spectra for hot Jupiter exoplanets. *Mon. Not. R. Astron. Soc.* **474**, 5158–5185 (2018).

84. Sing, D. K. Stellar limb-darkening coefficients for CoRot and Kepler. *Astron. Astrophys.* **510**, A21 (2010).

85. Polyansky, O. L. et al. ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *Mon. Not. R. Astron. Soc.* **480**, 2597–2608 (2018).

86. Huang, X., Gamache, R. R., Freedman, R. S., Schwenke, D. W. & Lee, T. J. Reliable infrared line lists for 13 co2 isotopologues up to e'=18,000cm-1 and 1500k, with line shape parameters. *J. Quant. Spectrosc. Ra.* **147**, 134–144 (2014).

87. Li, G. et al. Rovibrational line lists for nine isotopologues of the CO molecule in the X $^1\Sigma^+$ ground electronic state. *Astrophys. J. S.* **216**, 15 (2015).

88. Azzam, A. A. A., Tennyson, J., Yurchenko, S. N. & Naumenko, O. V. ExoMol molecular line lists – XVI. The rotation–vibration spectrum of hot H2S. *Mon. Not. R. Astron. Soc.* **460**, 4063–4074 (2016).

89. Hargreaves, R. J. et al. An accurate, extensive, and practical line list of methane for the HITEMP Database. *Astrophys. J. S.* **247**, 55 (2020).

90. Underwood, D. S. et al. ExoMol molecular line lists - XIV. The rotation-vibration spectrum of hot SO_2. *Mon. Not. R. Astron. Soc.* **459**, 3890–3899 (2016).

91. Allard, N. F., Spiegelman, F., Leininger, T. & Molliere, P. New study of the line profiles of sodium perturbed by H_2. *Astron. Astrophys.* **628**, A120 (2019).

92. Allard, N. F., Spiegelman, F. & Kielkopf, J. F. Study of the K-H{ 2} quasi-molecular line satellite in the potassium resonance line. *Astron. Astrophys.* **465**, 1085–1091 (2007).

Acknowledgements

This work is based on observations made with the NASA/ESA/CSA JWST. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract no. NAS 5-03127 for JWST. These observations are associated with program no. JWST-ERS-01366. Support for program no. JWST-ERS-01366 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract no. NAS 5-03127. This work benefited from the 2022 Exoplanet Summer Program in the Other Worlds Laboratory at the University of California, Santa Cruz, a program supported by the Heising-Simons Foundation. We thank E. Agol for constructive comments.

Author information

Authors and Affiliations

1. Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA

   Z. Rustamkulov & D. K. Sing

2. Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA

   D. K. Sing, N. H. Allen & L. Ramos-Rosado

3. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

   S. Mukherjee, A. L. Carter, N. M. Batalha, J. J. Fortney & K. Ohno

4. Johns Hopkins APL, Laurel, MD, USA

   E. M. May, K. B. Stevenson & J. Lustig-Yaeger

5. Center for Astrophysics, Harvard and Smithsonian, Cambridge, MA, USA

   J. Kirk & M. López-Morales

6. Department of Physics, Imperial College London, London, UK

   J. Kirk

7. Steward Observatory, University of Arizona, Tucson, AZ, USA

   E. Schlawin & M. Mansfield

8. School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

   M. R. Line, E. L. Shkolnik & L. Welbanks

9. Institute of Research on Exoplanets, Department of Physics, University of Montreal, Montreal, Québec, Canada

C. Piaulet, B. Benneke, P. Roy & M. Radica

10. NASA Ames Research Center, Moffett Field, CA, USA

N. E. Batalha

11. School of Earth and Planetary Sciences, National Institute of Science Education and Research (NISER), HBNI, Jatani, India

J. M. Goyal

12. Department of Physics, Utah Valley University, Orem, UT, USA

J. D. Lothringer

13. Department of Astronomy, University of Michigan, Ann Arbor, MI, USA

R. J. MacDonald

14. Department of Astronomy and Carl Sagan Institute, Cornell University, Ithaca, NY, USA

R. J. MacDonald, J. D. Turner, L. Flagg & Y. Hong

15. Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA

S. E. Moran

16. School of Physics, University of Bristol, HH Wills Physics Laboratory, Bristol, UK

H. R. Wakeford & L. Alderson

17. Space Telescope Science Institute, Baltimore, MD, USA

N. Espinoza, N. K. Nikolov & Q. Changeat

18. Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA

J. L. Bean & A. D. Feinstein

19. Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, USA

Z. K. Berta-Thompson

20. Department of Physics and Astronomy, University of Kansas, Lawrence, KS, USA

I. J. M. Crossfield & J. Brande

21. Earth and Planets Laboratory, Carnegie Institution of Washington, Washington, DC, USA

P. Gao, M. K. Alam & J. Teske

22. Max Planck Institute for Astronomy, Heidelberg, Germany

L. Kreidberg, T. Mikal-Evans, L. Mancini & S. Zieba

23. Harvard and Smithsonian, Center for Astrophysics, Cambridge, MA, USA

D. K. Powell

24. INAF - Astrophysics Observatory at Turin, Turin, Italy

P. E. Cubillos & L. Mancini

25. School of Physics, Trinity College Dublin, Dublin, Ireland

N. P. Gibson

26. Laboratory of Astrophysics at Bordeaux, CNRS, University of Bordeaux, Pessac, France

J. Leconte

27. University Observatory Munich, Ludwig Maximilian University, Munich, Germany

K. Molaverdikhani & K. Heng

28. Exzellenzcluster Origins, Garching, Germany

K. Molaverdikhani

29. University of the Côte d’Azur, Observatory of the Côte d’Azur, CNRS, Lagrange Laboratory, Nice, France

V. Parmentier

30. Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

V. Parmentier

31. Department of Physics, University of Oxford, Oxford, UK

J. Taylor

32. Centre for Exoplanets and Habitability, University of Warwick, Coventry, UK

P. J. Wheatley & E. Ahrer

33. Department of Physics, University of Warwick, Coventry, UK

P. J. Wheatley & E. Ahrer

34. Indian Institute of Technology, Indore, Indore, India

K. Aggarwal

35. School of Physical Sciences, The Open University, Milton Keynes, UK

A. Banerjee, J. K. Barstow & A. D. Feinstein

36. Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, the Netherlands

S. Barat & J. Désert

37. Centre for Astrobiology (CSIC-INTA), European Space Astronomy Centre Campus, University of Maria de Maeztu, Madrid, Spain

D. Barrado & J. Lillo-Box

38. BAER Institute, NASA Ames Research Center, Moffet Field, Mountain View, CA, USA

T. J. Bell

39. New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

J. Blecic

40. Center for Astro, Particle and Planetary Physics (CAP3), New York University Abu Dhabi, Abu Dhabi, UAE

J. Blecic

41. School of Physics and Astronomy, University of Leicester, Leicester, UK

S. Casewell

42. European Space Agency (ESA), ESA Baltimore Office, Baltimore, MD, USA

Q. Changeat

43. Department of Physics and Astronomy, University College London, London, UK

Q. Changeat

44. Centre for Exoplanet Science, University of St Andrews, St Andrews, UK

K. L. Chubb

45. Leiden Observatory, Leiden University, Leiden, the Netherlands

N. Crouzet, A. Louca, Y. Miguel & S. Zieba

46. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

T. Daylan

47. Department of Physics and Astronomy, KU Leuven, Leuven, Belgium

L. Decin

48. Planetary Science Group, Department of Physics and Florida Space Institute, University of Central Florida, Orlando, FL, USA

J. Harrington

49. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

R. Hu & T. Kataria

50. Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

R. Hu

51. Institute for Astrophysics, University of Vienna, Vienna, Austria

N. Iro

52. Department of Astronomy, University of Maryland, College Park, MD, USA

E. M.-R. Kempton

53. California Institute of Technology, IPAC, Pasadena, CA, USA

J. Krick

54. Department of Astronomy, University of Geneva, Geneva, Switzerland

M. Lendl & D. J. M. Petit dit de la Roche

55. Department of Physics, University of Rome 'Tor Vergata', Rome, Italy

L. Mancini

56. Department of Physics and Astronomy, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK

N. J. Mayne

57. SRON Netherlands Institute for Space Research, Leiden, the Netherlands

Y. Miguel

58. Institute for Astrophysics of Canarias (IAC), La Laguna, Tenerife, Spain

G. Morello & E. Palle

59. Department of Astrophysics, University of La Laguna, La Laguna, Tenerife, Spain

G. Morello

60. INAF Äì Palermo Astronomical Observatory, Palermo, Italy

G. Morello

61. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

B. V. Rackham

62. Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA

B. V. Rackham

63. Astronomy Department and Van Vleck Observatory, Wesleyan University, Middletown, CT, USA

S. Redfield

64. Institute of Astronomy, University of Cambridge, Cambridgeshire, UK

L. K. Rogers

65. Astrophysics Group, Keele University, Staffordshire, UK

J. Southworth

66. UVSQ, CNRS, CEA, House of Simulation, University of Paris-Saclay, Gif-sur-Yvette, France

P. Tremblin

67. Department of Physics, Brown University, Providence, RI, USA

G. S. Tucker

68. University of the City of Paris and University of Paris Est Creteil, CNRS, LISA, Paris, France

O. Venot

69. Astrophysics and Planetary Sciences, University of Colorado Boulder, Boulder, CO, USA

W. C. Waalkes

70. Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, CA, USA

X. Zhang

Contributions

All authors played a significant role in one or more of the following: development of the original proposal, management of the project, definition of the target list and observation plan, analysis of the data, theoretical modelling and preparation of this paper. Some specific contributions are listed as follows. N.E.B., J.L.B., Z.K.B-T., I.J.M.C., J.D., L.K., M.R.L., D.K.S., and K.B.S. provided overall program leadership and management. D.K.S., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., K.B.S., L.K., M.L.-M., M.R.L., N.M.B., V.P. and Z.K.B.-T. made significant contributions to the design of the program. K.B.S. generated the observing plan with input from the team. B.B., E.M.R.K., H.R.W., I.J.M.C., J.L.B., L.K., M.L.M., M.R.L., N.M.B. and Z.K.B.-T. led or coled working groups and/or contributed to significant strategic planning efforts such as the design and implementation of the prelaunch Data Challenges. A.L.C., D.K.S., E.S., N.E., N.P.G., and V.P. generated simulated data for prelaunch testing of methods. Z.R., D.K.S., J.K., E.S., E.M.M. and A.L.C. reduced the data, modelled the light curves and produced the planetary spectrum. S.M., M.R.L., J.M.G. and J.L. generated theoretical model grids for comparison with the data. Z.R., D.K.S., C.P., E.M.M., E.S., J.K., M.L. and S.M. made significant contributions to the writing of this paper. Z.R., D.K.S., S.M., E.M.M. and M.R.L. generated figures for this paper. B.V.R., D.K.P., I.J.M.C., J.T.,

J.M.G., J.L.B., M.L.M., R.H., R.J.M., S.E.M. and T.D. contributed to the writing of this paper.

Corresponding author

Correspondence to Z. Rustamkulov.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Drake Deming and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 A comparison of the extracted 1D spectrophotometry across the four reductions.

Plotted is the spectrophotometry with time on the x-axis and wavelength on the y-axis, with color indicating the relative flux. The transit is visible as a

dark band in the middle of the observation. All four reductions show nearly identical noise properties.

Extended Data Fig. 2 Demonstration of the impact of saturation.

Shown are the group-level median frames from the uncalibrated data products across the entire integration. The dashed blue line represents the empirically derived saturation level, with the orange dotted line representing 85% saturation, the level adopted in the Eureka! reduction. Grey shaded regions represent columns that reach 85% full well in a given group.

Extended Data Fig. 3 The wavelength-dependent central transit time in seconds.

Structure is apparent–the prominent water and carbon dioxide absorption features at 2.7 μm and 4.2 μm, respectively, appear to arrive ∼20 s after the optical continuum. A slope is also apparent from the blue side to the red. The error bars are 1-σ standard deviations.

Extended Data Fig. 4 A summary of the positional shifts of the trace, the wavelength-dependent light curve scatter, and the transit depth noise.

(**Top**) The X- and Y-shift vectors as measured by 1D cross correlation with FIREFLy. (**Middle**) The residual spectrophotometric light curves are shown for four representative spectral channels spanning the PRISM wavelength range with no temporal binning. The residual scatter is approximately Gaussian for each, as indicated by the histogram on the right y-axis. We validate this by performing Anderson-Darling tests on the residuals of the spectral and white-light curves, and find that all of the Anderson-Darling test statistics lie below the respective critical values 1% significance level. Therefore, we find that there is not sufficient evidence that the residuals are not normally distributed. (**Bottom**) The top two purple curves show the expected and measured normalised light curve root mean square (RMS)

residuals, with no temporal binning. Longward of 2 μm, the scatter in each light curve matches well with the expected noise as estimated by the jwst pipeline, which is dominated by photon noise. This agreement indicates the majority of the light curves reach near the photon limit. The transit depth uncertainties are also plotted below, including the white noise (blue, σ_w), red noise (red, σ_{red}), and total noise components (grey, σ_{tot}). Some wavelength bins have enhanced red noise, but the majority of the transmission spectrum is consistent with minimal red noise from residual systematic errors. The wavelengths affected by detector saturation are indicated by the grey shaded bar, with darker colors corresponding to quicker saturation. The colored dots are the measured RMS values from the light curves shown in the top panel.

Extended Data Fig. 5 Empirically derived stellar limb darkening coefficients fit with a quadratic law.

a, the fit u_+ coefficients (black) along with the theoretically predicted values derived from a 3D stellar model (red). The theoretical u_+ values with a constant offset of -0.065 ± 0.022 (purple) is also shown. The theoretical models predict the wavelength-to-wavelength shape of u_+ well. As u_+ is directly related to the intensity of the star at the stellar limb ref. [84], these findings suggest WASP-39A is 6% brighter at the limb than models predict. **b**, similar as **a**, but for the u_- coefficient. As the shape of the derived coefficients differs from the model prediction, u_- was left free to vary in the transmission spectral fits. The error bars are 1-σ standard deviations.

Extended Data Fig. 6 Comparison of the JWST NIRSpec PRISM data (black) to HST and VLT data from ref. 6,15 and WHT data from ref. [16], respectively.

The JWST spectrum was derived with the limb darkening fixed to the same 3D stellar model as in[6] to aid comparisons. With fixed limb darkening, the JWST transmission spectrum has lower overall transit depths especially at optical wavelengths. The broadband spectrum from the two space telescopes compares well, including the amplitude of the 1.4 μm water

feature first observed by HST/WFC3 and the Na feature near 0.6 µm observed by HST/STIS. The error bars are 1-σ standard deviations.

Extended Data Fig. 7 Best-fit models from ScCHIMERA, PICASO 3.0, ATMO, and Phoenix 1D RCTE model grids for WASP-39b.

The FIREFLy reduction is overlaid in the top panel. The top left inset panel shows the data and the models between 0.5—1.2 µm. All these models prefer super-solar atmospheric metallicities and cloudy atmospheres for WASP-39 b. The C/O ratio estimated by these models lies in the range 0.6–0.7. Additional SO_2 was injected in the PICASO 3.0 and ScCHIMERA grids to estimate the abundance of SO_2 required to explain the 4.0 µm feature, in a Bayesian framework. The ATMO and PHOENIX models are shown without any additionally injected SO_2 to emphasize that RCTE models do not predict such an SO_2 feature and chemical disequilibrium effects are required to explain the observed feature. The bottom panel shows the residuals from each best-fit model divided by the noise in the transit depth as a function of wavelength. The error bars are 1-σ standard deviations.

Extended Data Fig. 8 Each panel shows the residual spectrum of a particular gas.

This residual spectrum was obtained by removing one gas at a time from the best-fit model atmosphere and subtracting the recalculated model transmission spectrum without that gas from the data. This residual spectrum was then fitted with a Gaussian distribution (and a Voigt profile for Na) and a constant offset, in a Bayesian framework. The median fit (solid lines) along with the 1σ and 2σ confidence intervals are shown with shaded red and blue regions for the Gaussian fits and the constant offset fits, respectively. The Bayes factor between the two functional fits was used to determine the detection significance of each gas. Note that the wavelength range covered in each panel is different. The error bars are 1-σ standard deviations.

Extended Data Fig. 9 Models of varying metallicity (top) and C/O ratio (bottom) compared to the FIREFLy reduction.

A comparison of cloud-free PICASO 3.0 RCTE models across a span of metallicities with the best-fit C/O ratio (0.68) is shown in the top panel. Each line coloured from faded to deep pink represents models with different metallicities between sub-solar to super-solar values. The simultaneous lack of a prominent CH_4 feature at 2.3 and 3.3 μm and the presence of a strong CO_2 feature indicate that the observations disfavor a low-metallicity atmosphere. The bottom panel shows transmission spectrum models with different C/O ratios from sub-solar to super-solar values at 10×solar metallicity compared with the observed spectrum. The cloudy best-fit model obtained with the grid retrieval framework also has been shown in both the panels with the grey line. As before, the errorbars are 1σ standard deviations.

Extended Data Fig. 10 The wavelength-dependent contribution function.

The shaded regions highlight the parts of the atmosphere probed by the observed transmission data as a function of wavelength, as calculated from the best-fit model. This shows that the data mostly probe pressure ranges between 0.1 to 2 mbars. The CO_2 feature shows contribution at pressures approaching a microbar. The various shaded lines in pink show the volume mixing ratio of CH_4 (upper x-axis), from thermochemical equilibrium models, with different atmospheric metallicities at the best-fit C/O ratio of 0.68.

Extended Data Table 1 Best-fit orbital parameters as measured from the FIREFLy white light curve

Extended Data Table 2 An overview of the analysis procedures used by the independent data reductions

Extended Data Table 3 Overview of the best-fit model parameters obtained from each grid

Extended Data Table 4 Detection significances of individual opacity sources with our two techniques: Bayes factor analysis with gas

removal, and Gaussian/Voigt fits to the residual absorption profiles[85,86,87,88,89,90,91,92]

Supplementary information

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Rustamkulov, Z., Sing, D.K., Mukherjee, S. *et al.* Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM. *Nature* **614**, 659–663 (2023). https://doi.org/10.1038/s41586-022-05677-y

- Received: 22 September 2022
- Accepted: 21 December 2022

- Published: 09 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05677-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H**
  - Lili Alderson
  - Hannah R. Wakeford
  - Xi Zhang

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRCam**
  - Eva-Maria Ahrer
  - Kevin B. Stevenson
  - Xi Zhang

  Nature (2023)

- ## Early Release Science of the exoplanet WASP-39b with JWST NIRISS

    - Adina D. Feinstein
    - Michael Radica
    - Xi Zhang

    Nature (2023)

JWST opens a window on exoplanet skies

- Julia V. Seidel
- Louise D. Nielsen
- Subhajit Sarkar

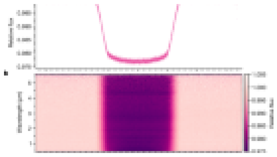
Nature News & Views Forum 15 Feb 2023

A population of red candidate massive galaxies ~600 Myr after the Big Bang

- Ivo Labbé
- Pieter van Dokkum
- Bingjie Wang

Nature Article 22 Feb 2023

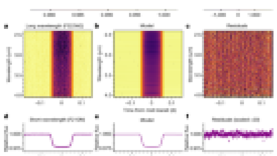
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05677-y



Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM

09 January 2023

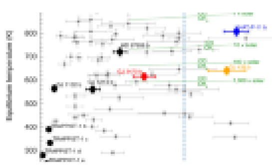
Z. Rustamkulov, D. K. Sing, … S. Zieba



Early Release Science of the exoplanet WASP-39b with JWST NIRCam

09 January 2023

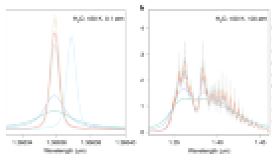
Eva-Maria Ahrer, Kevin B. Stevenson, … Xi Zhang



A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds

01 July 2019

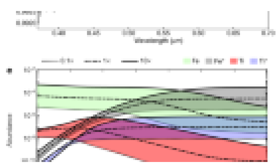
Björn Benneke, Heather A. Knutson, … Joshua Kammer



[The impending opacity challenge in exoplanet atmospheric characterization]

15 September 2022

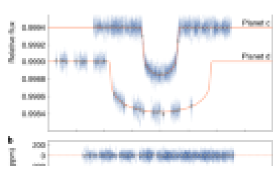
Prajwal Niraula, Julien de Wit, … Roman V. Kochanov



[Atomic iron and titanium in the atmosphere of the exoplanet KELT-9b]

15 August 2018

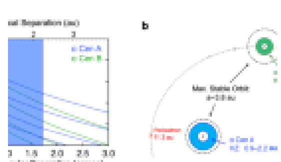
H. Jens Hoeijmakers, David Ehrenreich, … Luca Di Fabrizio



[Transit detection of the long-period volatile-rich super-Earth v2 Lupi d with CHEOPS]

28 June 2021

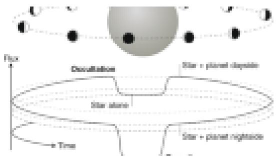
Laetitia Delrez, David Ehrenreich, … Nicholas A. Walton



Imaging low-mass planets within the habitable zone of α Centauri

10 February 2021

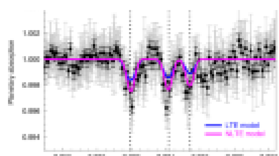
K. Wagner, A. Boehle, … T. de Zeeuw



Highlights of exoplanetary science from Spitzer

14 May 2020

Drake Deming & Heather A. Knutson



High-resolution detection of neutral oxygen and non-LTE effects in the atmosphere of KELT-9b

22 December 2021

Francesco Borsa, Luca Fossati, … Denis Shulyak

- Article
- Open Access
- Published: 09 January 2023

Early Release Science of the exoplanet WASP-39b with JWST

NIRSpec G395H

- Lili Alderson ,
- Hannah R. Wakeford ,
- Munazza K. Alam,
- Natasha E. Batalha,
- Joshua D. Lothringer,
- Jea Adams Redai,
- Saugata Barat,
- Jonathan Brande,
- Mario Damiano,
- Tansu Daylan,
- Néstor Espinoza,
- Laura Flagg,
- Jayesh M. Goyal,
- David Grant,
- Renyu Hu,
- Julie Inglis,
- Elspeth K. H. Lee,
- Thomas Mikal-Evans,
- Lakeisha Ramos-Rosado,
- Pierre-Alexis Roy,
- Nicole L. Wallack,
- Natalie M. Batalha,
- Jacob L. Bean,
- Björn Benneke,
- Zachory K. Berta-Thompson,
- Aarynn L. Carter,
- Quentin Changeat,
- Knicole D. Colón,
- Ian J. M. Crossfield,
- Jean-Michel Désert,
- Daniel Foreman-Mackey,
- Neale P. Gibson,
- Laura Kreidberg,
- Michael R. Line,

- Mercedes López-Morales,
- Karan Molaverdikhani,
- Sarah E. Moran,
- Giuseppe Morello,
- Julianne I. Moses,
- Sagnick Mukherjee,
- Everett Schlawin,
- David K. Sing,
- Kevin B. Stevenson,
- Jake Taylor,
- Keshav Aggarwal,
- Eva-Maria Ahrer,
- Natalie H. Allen,
- Joanna K. Barstow,
- Taylor J. Bell,
- Jasmina Blecic,
- Sarah L. Casewell,
- Katy L. Chubb,
- Nicolas Crouzet,
- Patricio E. Cubillos,
- Leen Decin,
- Adina D. Feinstein,
- Joanthan J. Fortney,
- Joseph Harrington,
- Kevin Heng,
- Nicolas Iro,
- Eliza M.-R. Kempton,
- James Kirk,
- Heather A. Knutson,
- Jessica Krick,
- Jérémy Leconte,
- Monika Lendl,
- Ryan J. MacDonald,
- Luigi Mancini,
- Megan Mansfield,
- Erin M. May,
- Nathan J. Mayne,

- Yamila Miguel,
- Nikolay K. Nikolov,
- Kazumasa Ohno,
- Enric Palle,
- Vivien Parmentier,
- Dominique J. M. Petit dit de la Roche,
- Caroline Piaulet,
- Diana Powell,
- Benjamin V. Rackham,
- Seth Redfield,
- Laura K. Rogers,
- Zafar Rustamkulov,
- Xianyu Tan,
- P. Tremblin,
- Shang-Min Tsai,
- Jake D. Turner,
- Miguel de Val-Borro,
- Olivia Venot,
- Luis Welbanks,
- Peter J. Wheatley &
- …
- Xi Zhang

Show authors

Nature **volume 614**, pages 664–669 (2023)

- 2411 Accesses

- 4 Citations

- 26 Altmetric

- Metrics details

Abstract

Measuring the abundances of carbon and oxygen in exoplanet atmospheres is considered a crucial avenue for unlocking the formation and evolution of exoplanetary systems[1,2]. Access to the chemical inventory of an exoplanet requires high-precision observations, often inferred from individual molecular detections with low-resolution space-based[3,4,5] and high-resolution ground-based[6,7,8] facilities. Here we report the medium-resolution ($R \approx 600$) transmission spectrum of an exoplanet atmosphere between 3 and 5 μm covering several absorption features for the Saturn-mass exoplanet WASP-39b (ref. [9]), obtained with the Near Infrared Spectrograph (NIRSpec) G395H grating of JWST. Our observations achieve 1.46 times photon precision, providing an average transit depth uncertainty of 221 ppm per spectroscopic bin, and present minimal impacts from systematic effects. We detect significant absorption from CO_2 (28.5σ) and H_2O (21.5σ), and identify SO_2 as the source of absorption at 4.1 μm (4.8σ). Best-fit atmospheric models range between 3 and 10 times solar metallicity, with sub-solar to solar C/O ratios. These results, including the detection of SO_2, underscore the importance of characterizing the chemistry in exoplanet atmospheres and showcase NIRSpec G395H as an excellent mode for time-series observations over this critical wavelength range[10].

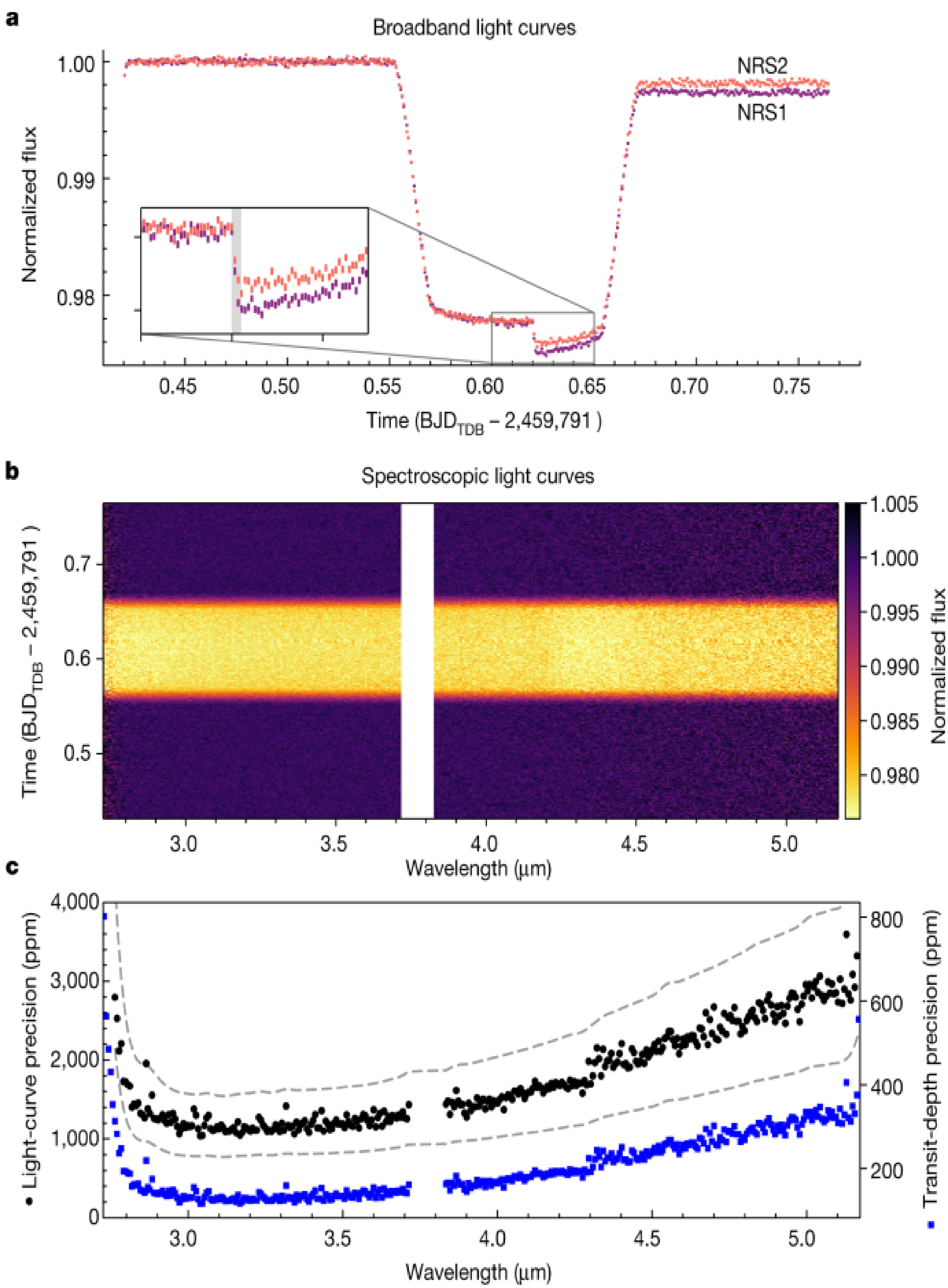
Main

We obtained a single-transit observation of WASP-39b using the NIRSpec[11,12] G395H grating on 30–31 July 2022 between 21:45 and 06:21 UTC using the Bright Object Time Series mode. WASP-39b is a hot (T_{eq} = 1,120 K), low-density giant planet with an extended atmosphere. Previous spectroscopic observations have shown prominent atmospheric absorption by Na, K and H_2O (refs. [3,4,13,14,15]), with tentative evidence of CO_2 from infrared photometry[4]. Atmospheric models fitted to the spectrum have inferred metallicities (amount of heavy elements relative to the host star) from 0.003 to 300 times solar[3,15,16,17,18,19,20], which makes it difficult to ascertain the formation pathway of the planet[21,22]. The host, WASP-39, is a G8-type star that shows little photometric variability[23] and has nearly solar elemental abundance patterns[24]. The quiet host and extended planetary atmosphere make WASP-39b an ideal exoplanet for transmission

spectroscopy[25]. The transmission spectrum of WASP-39b was observed as part of the JWST Transiting Exoplanet Community Director's Discretionary Early Release Science (JTEC ERS) Program[26,27] (ERS-1366; principal investigators Natalie M. Batalha, Jacob L. Bean and Kevin B. Stevenson), which uses four instrument configurations to test their capabilities and provide lessons learned for the community.

The NIRSpec G395H data were recorded with the 1.6″ × 1.6″ fixed slit aperture using the SUB2048 subarray and NRSRAPID readout pattern, with spectra dispersed across both the NRS1 and NRS2 detectors. Over the roughly 8-h duration of the observation, a total of 465 integrations were taken, centred around the 2.8-h transit. We obtained 70 groups per integration, resulting in an effective integration time of 63.14 s. During the observation, the telescope experienced a 'tilt event', a spontaneous and abrupt change in the position of one or more mirror segments, causing changes in the point spread function (PSF) and hence jumps in flux[28]. The tilt event occurred mid-transit, affecting approximately three integrations and resulted in a noticeable step in the flux time series, the size of which is dependent on wavelength (Fig. 1 and Methods). The tilt event also affects the PSF, with the full width at half maximum (FWHM) of the spectral trace showing a step-function-like shape (see Extended Data Figs. 2 and 3).

Fig. 1: Light-curve precisions achieved for WASP-39b with NIRSpec G395H.



a, Raw, uncorrected broadband light curves from the NRS1 (purple) and NRS2 (red) detectors, demonstrating the lack of dominant systematic trends

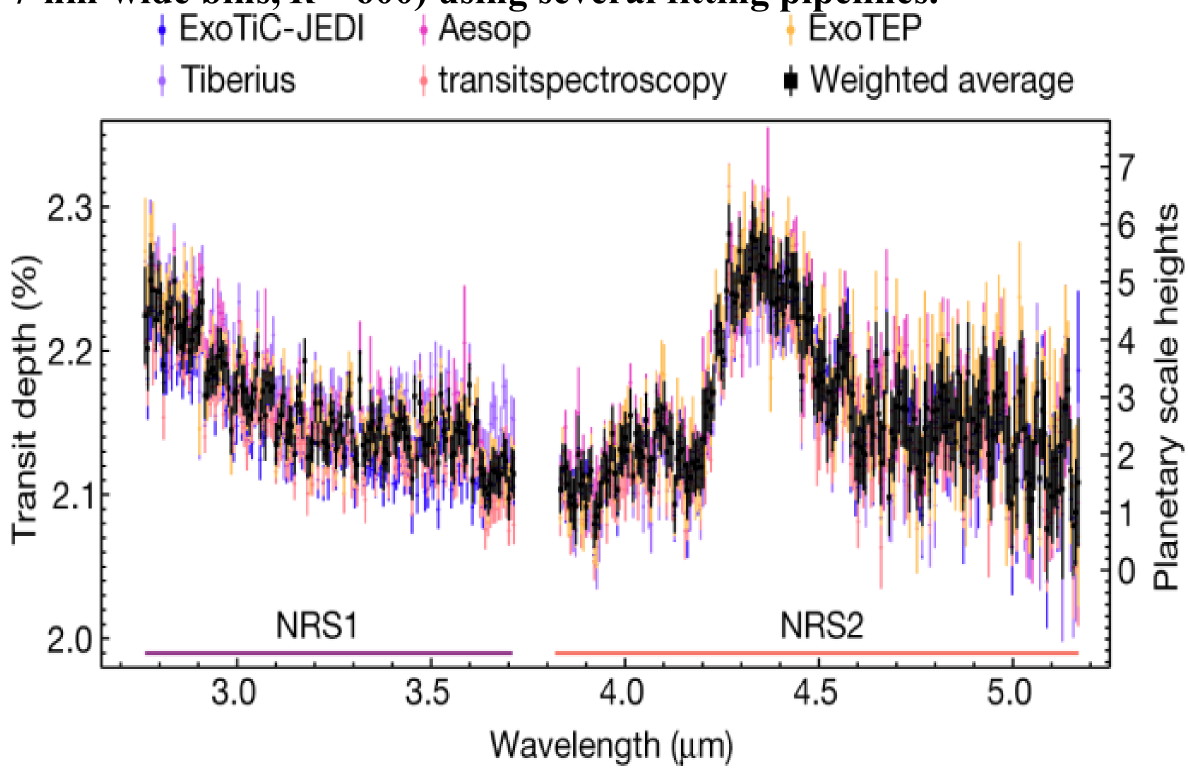
in the light curves. The inset shows the drop in flux (grey-shaded region) caused by a mirror-tilt event, resulting in a distinct change in flux between NRS1 and NRS2 after the tilt event (see Extended Data Figs. 2 and 3). **b**, Pixel intensity map of the spectroscopic light curves after correction for the tilt event and further instrument systematics. **c**, Light-curve precision obtained per spectroscopic bin (black) compared with 1 and 2 times photon noise expectations (grey dashed lines) and the measured precision on the transit depth (blue). The gap between the two detectors (3.72–3.82 µm) is highlighted in the middle and bottom plots. All data shown are from fitting pipeline 1 (see Methods).

We produced several reductions of the observations using independent analysis pipelines (see Methods). For each reduction, we created broadband and spectroscopic light curves in the ranges 2.725–3.716 µm for NRS1 and 3.829–5.172 µm for NRS2 using 10-pixel-wide bins (≈0.007 µm, median resolution $R \approx 600$), excluding the detector gap between 3.717–3.823 µm. The light curves show a settling ramp during the first ten integrations (≈631.4 s), with a linear slope across the entire observation for NRS1. We otherwise see no substantial systematic trends and achieve improvements in precision from raw uncorrected to fitted broadband light curves of 1.63 to 1.03 times photon noise for NRS1 and 1.95 to 1.31 times for NRS2. The flux jump caused by the mirror-tilt event could be corrected by detrending against the spectral trace x and y positional shifts, normalizing the light curves or fitting the light curves with a step function (see Methods). We produced several fits from each set of light curves, resulting in a total of 11 independently measured transmission spectra. Figure 1 demonstrates that our spectroscopic light curves achieve precisions close to photon noise, with a median precision of 1.46 times photon noise across the full wavelength range (see Extended Data Fig. 4).

We show transmission spectra from several combinations of independent reductions and light-curve-fitting routines in Fig. 2, along with the weighted average of all 11 transmission spectra with the unweighted mean uncertainty produced by our analyses (see Methods). We find that using different combinations of reduction and fitting methods results in consistent transmission spectra (see Methods and Extended Data Fig. 5). Although we see some artefacts at the edges of the detectors (see Fig. 3, bottom panel) that may be caused by uncharacterized systematics, these only affect a small

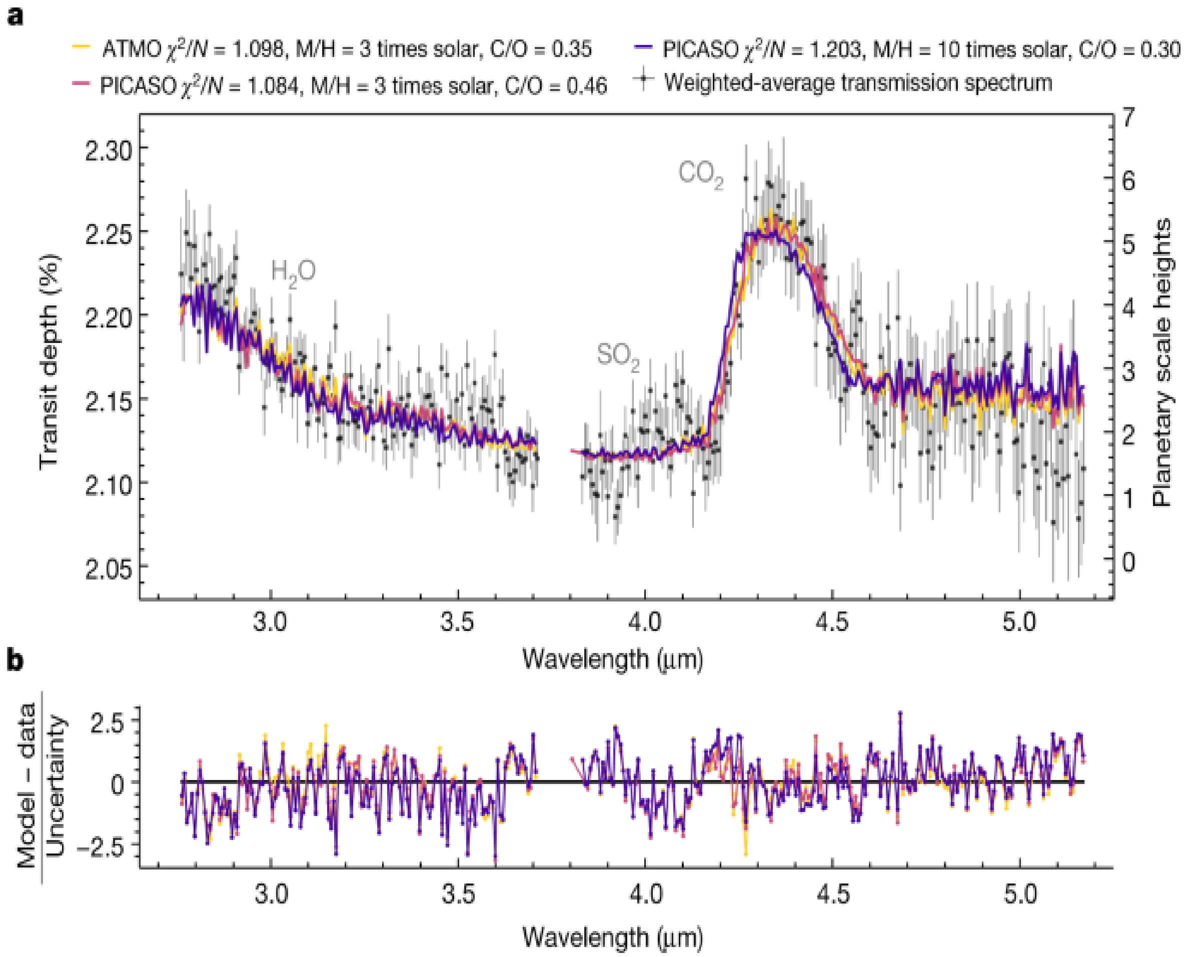
number of wavelength bins. Our resulting averaged NIRSpec G395H spectrum shows increased absorption towards bluer wavelengths short of 3.7 μm and a prominent absorption feature between 4.2 and 4.5 μm, along with a smaller-amplitude absorption feature at 4.1 μm and a narrow feature around 4.56 μm.

Fig. 2: WASP-39b transmission spectra measured at 10-pixel resolution (≈7-nm-wide bins, *R* ≈ 600) using several fitting pipelines.



We show the resultant spectra from five out of 11 independent fitting pipelines, which used distinct analysis methods to demonstrate the robust structure of the spectrum (see Methods for details on each fitting pipeline and comparative statistics). The black points show the weighted-average transmission spectrum computed from the transit depth values in each bin weighted by $1/\sigma^2$, in which σ is the uncertainty on the data point from each of the 11 fitting pipelines. The error bars were computed from the unweighted mean uncertainty in each bin (see Extended Data Fig. 5). All spectra show consistent broadband absorption short of 3.7 μm, around 4.1 μm and from 4.2 to 4.5 μm.

Fig. 3: Spectra from three independent 1D RCTE models and their residuals, fit to the weighted-average WASP-39b G395H transmission spectrum.



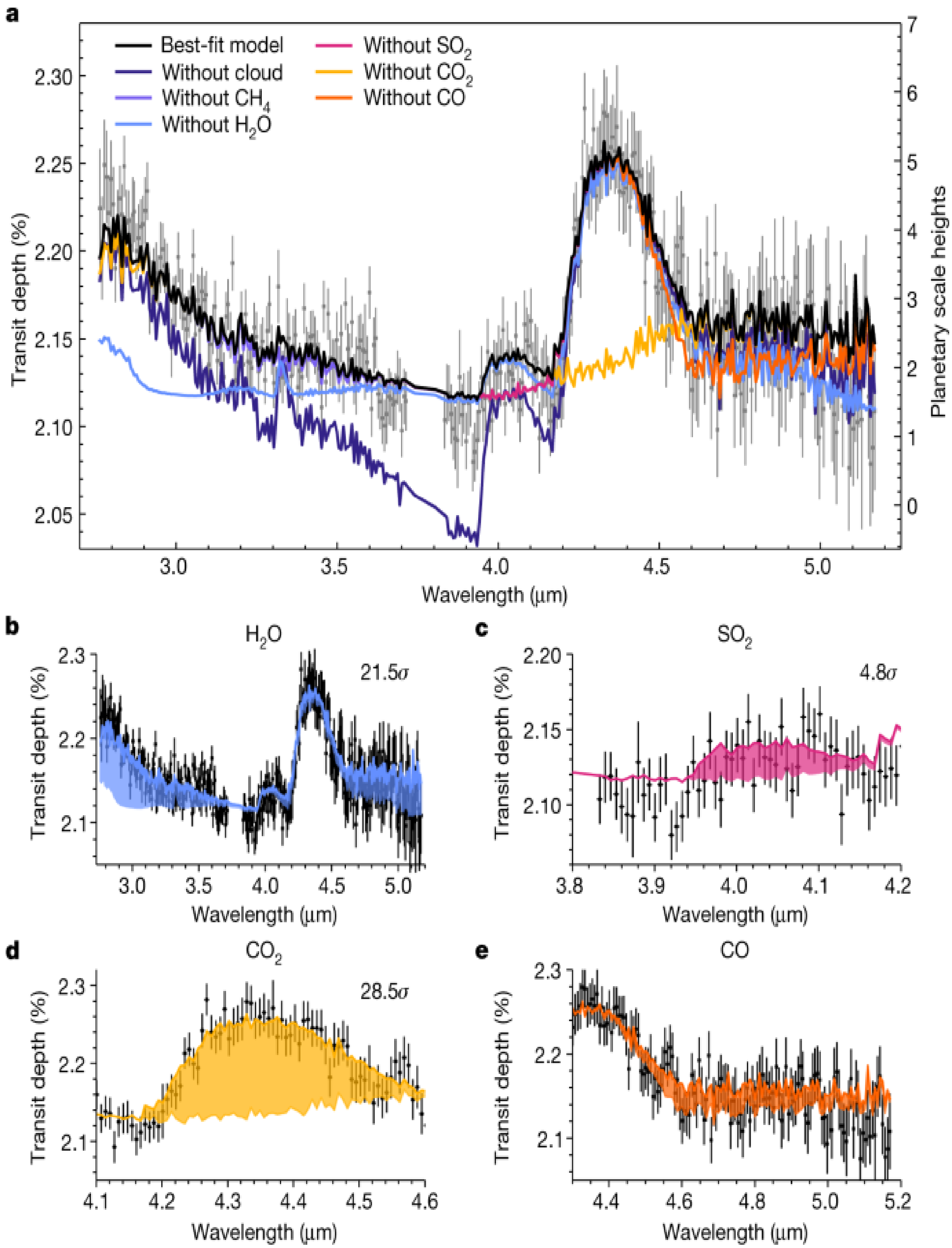
a, Spectra from the three models. **b**, Their residuals. The models are dominated by absorption from H_2O and CO_2 with a grey-cloud-top pressure corresponding to ≈1 mbar. The models find that the data are best explained by 3–10 times solar metallicity (M/H) and sub-solar to solar C/O (C/O = 0.30–0.46). The extra absorption owing to SO_2, seen in the spectrum around 4.1 µm, is not included in the RCTE model grids and causes a marked impact on the χ^2/N (see Fig. 4).

We compared the weighted-average G395H transmission spectrum to three grids of 1D radiative–convective–thermochemical equilibrium (RCTE) atmosphere models of WASP-39b (see Methods and Extended Data Table 2), containing a total of 10,308 model spectra. The best-fit models from each

grid provide a reduced chi-square per data point (χ^2/N) of 1.08–1.20 for our 344-data-point transmission spectrum (Fig. 3). The increased absorption at blue wavelengths across NRS1 is consistent with absorption from H_2O (at 21.5σ; see Methods), whereas the large bump in absorption between 4.2 and 4.5 μm (ref. [29]) can be attributed to CO_2 (28.5σ). H_2O and CO_2 are expected atmospheric constituents for near-solar atmospheric metallicities, with the CO_2 abundance increasing nonlinearly with higher metallicity[30]. The spectral feature at 4.56 μm (3.3σ) is unidentified at present but does not correlate with any obvious detector artefacts and is reproduced by several independent analyses. The absorption feature at 4.1 μm is also not seen in the RCTE model grids. After an exhaustive search for possible opacity sources (S.-M. Tsai et al., manuscript in preparation), described in the corresponding NIRSpec PRISM analysis[31], we interpret this feature as SO_2 (4.8σ), as it is the best candidate at this wavelength.

Although SO_2 would have volume mixing ratios (VMRs) of less than 10^{-10} throughout most of the observable atmosphere in thermochemical equilibrium, coupled photochemistry of H_2S and H_2O can produce SO_2 on giant exoplanets, with the resulting SO_2 mixing ratio expected to increase with increasing atmospheric metallicity[32,33,34]. We find that a VMR of approximately 10^{-6} of SO_2 is required to fit the spectral feature at 4.1 μm in the transmission spectrum of WASP-39b, consistent with lower-resolution NIRSpec PRISM observations of this planet[31] and previous photochemical modelling of super-solar metallicity giant exoplanets[34,35]. Figure 4 shows a breakdown of the contributing opacity sources for the lowest χ^2/N best-fit model (PICASO 3.0) with VMR = $10^{-5.6}$ injected SO_2. The inclusion of SO_2 in the models results in an improved χ^2/N and is detected at 4.8σ (see Methods), confirming its presence in the atmosphere of WASP-39b.

Fig. 4: Contribution of opacity sources to the best-fitting model with injected SO_2.



a, The lowest χ^2/N best-fitting model (PICASO in Fig. 3) with an injected abundance of $10^{-5.6}$ (VMR) SO_2. We also show this model with a selection

of the anticipated absorbing species and the cloud opacity removed to indicate their contributions to the model. The inclusion of SO_2 in the model decreases the χ^2/N from 1.08 (shown in Fig. 3) to 1.02, resulting in a 4.8σ detection (see Extended Data Table 3). **b**–**e**, The effect of removing the corresponding molecular opacity from the spectrum (shaded region). Our best-fit model is also affected by minor opacities from CO, H_2S, OCS and CH_4, although their spectral features cannot be robustly detected in the spectrum. We show a model without CO and CH_4 in **a** to demonstrate this, with the minor contribution by CO also highlighted in **e**.

We also look for evidence of CH_4, CO, H_2S and OCS (carbonyl sulfide) because their near-solar chemical equilibrium abundances could result in a contribution to the spectrum. We see no evidence of CH_4 in our spectrum between 3.0 and 3.6 μm (ref. [23]), which is indicative of C/O < 1 (ref. [36]) and/or photochemical destruction[34,37]. With regards to CO, H_2S and OCS, we were unable to conclusively confirm their presence with these data. In particular, CO, H_2O, OCS and our modelled cloud deck all have overlapping opacity, which creates a pseudo-continuum from 4.6 to 5.1 μm (see Figs. 3 and 4). Therefore, we were unable to unambiguously identify the individual contributions from CO and other molecules over this wavelength region at the resolution presented in this work.

Our models show an atmosphere enriched in heavy elements, with best-fit parameters ranging from 3 to 10 times solar metallicity, given the spacing of individual model grids (see Methods). The spectra also indicate C/O ratios ranging from sub-solar to solar depending on the grid used, informed by the relative strength of absorption from carbon-bearing molecules to oxygen-bearing molecules. The interpretation of the relatively high resolution and precision of the G395H spectrum seems to be sensitive to the treatment of aerosols in the model, with one grid preferring 3 times solar metallicity when using a wavelength-dependent cloud opacity and physically motivated vertical cloud distribution[38] but 10 times solar metallicity when assuming a grey cloud. In general, forward model grids fit the main features of the data but do not place statistically significant constraints on many of the atmospheric parameters (see Methods). Future interpretation of the JTEC ERS WASP-39b data with Bayesian retrieval analyses will provide robust

confidence intervals for these planetary properties and explore the degree to which these data are sensitive to modelling assumptions (for example, chemical equilibrium versus disequilibrium) and parameter degeneracies (for example, clouds versus atmospheric metallicity).

We are able to strongly rule out an absolute C/O ≥ 1 scenario ($\chi^2/N \geq 3.97$), which has previously been proposed for gas-dominated accretion at wide orbital radii beyond the CO_2 ice line at which the gas may be carbon-rich[39]. Our C/O upper limit, therefore, suggests that WASP-39b may have either formed at smaller orbital radii with gas-dominated accretion or that the accretion of solids enriched the atmosphere of WASP-39b with oxygen-bearing species[2]. The level of metal enrichment (3–10 times solar) is consistent with similar measurements of Jupiter and Saturn[40,41], potentially suggesting core-accretion formation scenarios[42], and is consistent with upper limits from interior-structure modelling[43]. These NIRSpec G395H transmission spectroscopy observations demonstrate the promise of robustly characterizing the atmospheric properties of exoplanets with JWST unburdened by substantial instrumental systematics, unravelling the nature and origins of exoplanetary systems.

Methods

Data reduction

We produced several analyses of stellar spectra from the Stage 1 2D spectral images produced using the default STScI JWST Calibration Pipeline[44] (‘rateints’ files) and by means of customized runs of the STScI JWST Calibration Pipeline with user-defined inputs and processes for steps such as the ‘jump detection’ and ‘bias subtraction’ steps.

Each pipeline starts with the raw ‘uncal’ 2D images that contain group-level products. As we noticed that the default superbias images were of poor quality, we produced two customized runs of the JWST Calibration Pipeline, using either the default bias step or a customized version. The customized step built a pseudo-bias image by computing the median pixel value in the first group across all integrations and then subtracted the new bias image

from all groups. We note that the poor quality of the default superbias images affects NRS1 more notably than NRS2, and this method could be revised once a better superbias is available.

Before ramp fitting, both our standard and custom bias step runs of the edited JWST Calibration Pipeline ‘destriped’ the group-level images to remove so-called ‘1/*f* noise’ (correlated noise arising from the electronics of the readout pattern, which appears as column striping in the subarray images[11,12]). Our group-level destriping step used a mask of the trace 15σ from the dispersion axis for all groups within an integration, ensuring that a consistent set of pixels is masked within a ramp. The median values of non-masked pixels in each column were then computed and subtracted for each group.

The results of our customized runs of the JWST Calibration Pipeline are a set of custom group-level destriped products and custom bias-subtracted group-level destriped products. In both cases, the ramp-jump detection threshold of the JWST Calibration Pipeline was set to 15σ (as opposed to the default of 4σ), as it produced the most consistent results at the integration level. In both custom runs of the JWST Calibration Pipeline, all other steps and inputs were left at the default values.

For all analyses, wavelength maps from the JWST Calibration Pipeline were used to produce wavelength solutions, verified against stellar absorption lines, for both detectors. The mid-integration times in BJD_{TDB} were extracted from the image headers for use in producing light curves. None of our data-reduction pipelines performed a flat-field correction, as the available flat fields were of poor quality and unexpectedly removed portions of the spectral trace. In general, we found that 1/*f* noise can be corrected at either the group or integration levels to similar effect; however, correction at the group level with a repeated column-by-column cleaning step at the integration level probably results in cleaner 1D stellar spectra. This was particularly true for NRS2, owing to the limited number of columns in which the unilluminated region on the detector extends both above and below the spectral trace, as shown in Extended Data Fig. 1.

Below we detail each of the independent data-reduction pipelines used to produce the time series of stellar spectra from our G395H observations.

ExoTiC-JEDI pipeline

We used the Exoplanet Timeseries Characterisation - JWST Extraction and Diagnostics Investigator (ExoTiC-JEDI[45]) pipeline on our custom group-level destriped products, treating each detector separately. Using the data-quality flags produced by the JWST Calibration Pipeline, we replaced any pixels identified as bad, saturated, dead, hot, low quantum efficiency or no gain value with the median value of surrounding pixels. We also searched each integration for pixels that were spatial outliers from the median of the surrounding 20 pixels in the same row by 6σ (to remove permanently affected 'bad' pixels) or outliers from the median of that pixel in the surrounding ten integrations in time by 20σ (to identify high-energy short-term effects such as cosmic rays) and replaced the outliers with the median values. To obtain the trace position and FWHM, we fitted a Gaussian to each column of an integration, finding a median standard deviation of 0.7 pixels. A fourth-order polynomial was fitted through the trace centres and the widths, which were smoothed with a median filter, to obtain a simple aperture region. This region extended 5 times the FWHM of the spectral trace, above and below the centre, corresponding to a median aperture width of 7 pixels. To remove any remaining $1/f$ and background noise from each integration, we subtracted the median of the unilluminated region in each column by masking all pixels that were 5 pixels away from the aperture. For each integration, the counts in each row and column of the aperture region were summed using an intrapixel extraction, taking the relevant fractional flux of the pixels at the edge of the aperture and cross-correlated to produce x-pixel and y-pixel shifts for detrending (see Extended Data Fig. 2). On average, the x-pixel shift represents movement of 1×10^{-4} and 8×10^{-6} of a pixel for NRS1 and NRS2, respectively. The aperture column sums resulted in 1D stellar spectra with errors calculated from photon noise after converting from data numbers using the gain factor. This reduction is denoted hereafter as ExoTiC-JEDI [V1].

We produced further 1D stellar spectra from both the custom group-level destriped product and custom bias-subtracted group-level destriped products using the ExoTiC-JEDI pipeline as described above, but with further cleaning by repeating the spatial outliers step. The reduction with further cleaning using the custom group-level destriped products is hence denoted

as ExoTiC-JEDI [V2] and the reduction with further cleaning using the custom bias-subtracted group-level destriped products is hence denoted as ExoTiC-JEDI [V3].

Tiberius pipeline

We used the Tiberius pipeline, which builds on the LRG-BEASTS spectral reduction and analysis pipelines[15,46,47], on our custom group-level destriped products. For each detector, we created bad-pixel masks by manually identifying hot pixels in the data. We then combined them with pixels flagged as greater than 3σ above the defined background. Before identifying the spectral trace, we interpolated each column of the detectors onto a grid 10 times finer than the initial spatial resolution. This step reduces the noise in the extracted data by improving the extraction of flux at the sub-pixel level, particularly where the edges of the photometric aperture bisect a pixel. We also interpolated over the bad pixels using their nearest-neighbouring pixels in x and y.

We traced the spectra by fitting Gaussians at each column and used a running median, calculated with a moving box with a width of five data points, to smooth the measured centres of the trace. We fitted these smoothed centres with a fourth-order polynomial, removed points that deviated from the median by 3σ and refitted with a fourth-order polynomial. To remove any residual background flux not captured by the group-level destriping, we fitted a linear polynomial along each column, masking the stellar spectrum. This was defined by an aperture with a width of 4 pixels centred on the trace. We also masked an extra 7 pixels on either side of the aperture so that the background was not fitting the wings of the stellar PSF and we clipped any pixels in the background that deviated by more than 3σ from the mean for that particular column and frame. After removing the background in each column, the stellar spectra were then extracted by summing within a 4-pixel-wide aperture and correcting for pixel oversampling caused by the interpolation onto a finer grid, as described above. The uncertainties in the stellar spectra were calculated from the photon noise before background subtraction.

transitspectroscopy pipeline

We used the transitspectroscopy pipeline[48] on the ‘rateints’ products of the JWST Calibration Pipeline, treating each detector separately. The trace position was found from the median integration by cross-correlating each column with a Gaussian function, removing outliers using a median filter with a 10-pixel-wide window and smoothing the trace with a spline. We removed $1/f$ noise from the ‘rateints’ products by masking all pixels within 10 pixels from the centre of the trace and calculating and removing the median value from all columns. We then used optimal extraction[49] to obtain the 1D stellar spectra, with a 5-pixel-wide aperture above and below the trace. This allowed us to treat bad pixels and cosmic rays that had not been accounted for or masked in the ‘rateints’ products in an automated fashion. To monitor systematic trends in the observations, we also calculated the trace centre as described above and the FWHM for all integrations. The FWHM was calculated at each column and at each integration by first subtracting each column to half the maximum value in it, with a spline used to interpolate the profile. The roots of this profile were then found to estimate the FWHM.

Eureka! pipeline

We used two customized versions of the Eureka! pipeline[50], which combines standard steps from the JWST Calibration Pipeline with an optimal extraction scheme to obtain the time series of stellar spectra.

The first Eureka! reduction used our custom group-level destriped products and applied Stages 2 and 3 of Eureka! Stage 2, a wrapper of the JWST Calibration Pipeline, followed the default settings up to the flat fielding and photometric calibration steps, which were both skipped. Stage 3 of Eureka! was then used to perform the background subtraction and extraction of the 1D stellar spectra. We started by correcting for the curvature of NIRSpec G395H spectra by shifting the detector columns by whole pixels, to bring the peak of the distribution of the counts in each column to the centre of our subarray. To ensure that this curvature correction was smooth, we computed the shifts in each column for each integration from the median integration frame in each segment and applied a running median to the shifts obtained for each column. The pixel shifts were applied with periodic boundary conditions, such that pixels shifted upwards from the top of the subarray

appeared at the bottom after the correction, ensuring no pixels were lost. We applied a column-by-column background subtraction by fitting and subtracting a flat line to each column of the curvature-corrected data frames, obtained by fitting all pixels further than six pixels from the central row. We also performed a double iteration of outlier rejection in time with a threshold of 10σ, along with a 3σ spatial outlier-rejection routine, to ensure that bad pixels were not biasing our background correction. These outlier-rejection thresholds were selected to remove clear outliers in the data and provide a balance with the background subtraction step. We performed optimal extraction using an extraction profile defined from the median frame, the central nine rows of our subarray (four rows on either side of the central row). We also measured the vertical shift in pixels of the spectrum from one integration to the other using cross-correlation and the average PSF width at each integration, obtained by adding all columns together and fitting a Gaussian to the profile to estimate its width. This reduction is henceforth denoted as Eureka! [V1].

The second Eureka! reduction (Eureka! [V2]) used the 'rateints' outputs of the JWST Calibration Pipeline and applied Stage 2 of Eureka! as described above, with a modified version of Stage 3. In this reduction, we corrected the curvature of the trace using a spline and found the centre of the trace using the median of each column. We removed $1/f$ noise by subtracting the mean from each column, excluding the region 6 pixels away from the trace, sigma-clipping outliers at 3σ. We extracted the 1D stellar spectra using a 4-pixel-wide aperture on either side of the trace centre.

Limb-darkening

Limb-darkening is a function of the physical structure of the star that results in variations in the specific intensity of the light from the centre of the star to the limb, such that the limb looks darker than the centre. This is because of the change in depth of the stellar atmosphere being observed. At the limb of the star, the region of the atmosphere being observed at slant geometry is at higher altitudes and lower density, and thus lower temperatures, compared with the deeper atmosphere observed at the centre of the star, at which hotter, denser layers are observed. The effect of limb-darkening is most prominent at shorter wavelengths, resulting in a more U-shaped light curve

compared with the flat-bottomed light curves observed at longer wavelengths. To account for the effects of limb-darkening on the time-series light curves, we used analytical approximations for computing the ratio of the mean intensity to the central intensity of the star. The most commonly used limb-darkening laws for exoplanet transit light curves are the quadratic, square-root and nonlinear four-parameter laws[51]:

Quadratic:

$$\frac{I(\mu)}{I(1)}=1-{u}_{1}\left(1-\mu \right)-{u}_{2}{\left(1-\mu \right)}^{2}$$

Square-root:

$$\frac{I(\mu)}{I(1)}=1-{s}_{1}\left(1-\mu \right)-{s}_{2}\left(1-\sqrt{\mu }\right)$$

Nonlinear four-parameter:

$$\frac{I(\mu)}{I(1)}=1-{c}_{1}(1-{\mu }^{0.5})-{c}_{2}(1-\mu)-{c}_{3}(1-{\mu }^{1.5})-{c}_{4}(1-{\mu }^{2})$$
(1)

in which $I(1)$ is the specific intensity in the centre of the disk, u_1, u_2, s_1, s_2, c_1, c_2, c_3 and c_4 are the limb-darkening coefficients and $\mu = \cos(\gamma)$, in which γ is the angle between the line of sight and the emergent intensity.

The limb-darkening coefficients depend on the particular stellar atmosphere and therefore vary from star to star. For consistency across all of the light-curve fitting, we used 3D stellar models[52] for $T_{eff} = 5{,}512$ K, $\log(g) = 4.47$ cgs and Fe/H = 0.0, along with the instrument throughput to determine I and μ. As instrument commissioning showed that the throughput was higher than the pre-mission expectations[53], a custom throughput was produced from the median of the measured ExoTiC-JEDI [V2] stellar spectra, divided by the stellar model and Gaussian smoothed.

For the limb-darkening coefficients that were held fixed, we used the values computed using the ExoTiC-LD[54,55] package, which can compute the linear,

quadratic and three-parameter and four-parameter nonlinear limb-darkening coefficients[51,56]. To compute and fit for the coefficients from the square-root law, we used previously outlined formalisms[57,58]. We highlight that we do not see any dependence in our transmission spectra on the limb-darkening procedure used across our independent reductions and analyses.

Light-curve fitting

From the time series of extracted 1D stellar spectra, we created our broadband transit light curves by summing the flux over 2.725–3.716 μm for NRS1 and 3.829–5.172 μm for NRS2. For the spectroscopic light curves, we used a common 10-pixel binning scheme within these wavelength ranges to generate a total of 349 spectroscopic bins (146 for NRS1 and 203 for NRS2). We also tested wider and narrower binning schemes but found that 10-pixel-wide bins achieved the best compromise between the noise in the spectrum and showcasing the abilities of G395H across analyses. In our analyses, we treated the NRS1 and NRS2 light curves independently to account for differing systematics across the two detectors. To construct the full NIRSpec G395H transmission spectrum of WASP-39b, we fitted the NRS1 and NRS2 broadband and spectroscopic light curves using 11 independent light-curve-fitting codes, which are detailed below. When starting values were required, all analyses used the same system parameters[37]. In many of our analyses, we detrended the raw broadband and spectroscopic light curves using the time-dependent decorrelation parameters for the change in the FWHM of the spectral trace or the shift in *x*-pixel and *y*-pixel positions (Extended Data Fig. 2). We also used various approaches to account for the mirror-tilt event, which we found to have a smaller effect at longer wavelengths (Extended Data Fig. 3).

Using fitting pipeline 1, we measured a centre of transit time (T_0) of T_0 = 2,459,791.612039 ± 0.000017 $\mathrm{BJD_{TDB}}$ and T_0 = 2,459,791.6120689 ± 0.000021 $\mathrm{BJD_{TDB}}$ computed from the NRS1 and NRS2 broadband light curves, respectively; most of the fitting pipelines obtained T_0 within 1σ of the quoted uncertainty.

For each of our analyses, we computed the expected photon noise from the raw counts taking into account the instrument read noise (16.18 e^- on NRS1

and 17.75 e^- on NRS2), gain (1.42 for NRS1 and 1.62 for NRS2) and the background counts (which are found to be negligible after cleaning) and compared it to the final signal-to-noise ratio in our light curves (see Fig. 1). We also determine the level of white and red noise in our spectroscopic light curves by computing the Allan deviation[59], which is used to measure the deviation from the expected photon noise by binning the data into successively smaller bins (that is, fewer data points per bin) and calculating the signal-to-noise ratio achieved[60]. Extended Data Fig. 4 shows the Allan deviation for three of the 11 reductions performed on the data (see the ExoTiC-ISM noise_calculator function[54]).

Although there is a general consensus across each of the data analyses, by comparing the results of each fitting pipeline, we were better able to evaluate the impact of different approaches to the data reduction, such as the removal of bad pixels. For future studies, we recommend the application of several pipelines that use differing analysis methods, such as the treatment of limb-darkening, systematic effects and noise removal. No single pipeline presented on its own can fully evaluate the measured impact of each effect, given the differing strategies, targets and potential for chance events such as a mirror tilt with each observation. In particular, attention should be paid to $1/f$ noise removal at the group versus integration levels for observations with fewer groups per integration than this study.

Below, we detail each of our 11 fitting pipelines and summarise them in Extended Data Table 1.

Fitting pipeline 1: ExoTiC-JEDI

We fitted the broadband and spectroscopic light curves produced from the ExoTIC-JEDI [V3] stellar spectra using the least-squares optimizer, scipy.optimize lm (ref. [61]). We simultaneously fitted a batman transit model[62] with a constant baseline and systematics models for data pre-tilt and post-tilt event, fixing the centre of transit time T_0, the ratio of the semi-major axis to stellar radius $a/R_\star$ and the inclination i to the broadband light-curve best-fit values. The systematics models included a linear regression on x and y, for which x and y are the measured trace positions in the dispersion and

cross-dispersion directions, respectively. We accounted for the tilt event by normalizing the light curve pre-tilt by the median pre-transit flux and normalizing the light curve post-tilt by the median post-transit flux. We discarded the first 15 integrations and the three integrations during the tilt event. Fourteen-pixel columns were discarded owing to outlier pixels directly on the trace. We fixed the limb-darkening coefficients to the four-parameter nonlinear law.

Fitting pipeline 2: Tiberius

We used the broadband light curves generated from the Tiberius stellar spectra and fitted for the ratio of the planet to stellar radii $R_p/R_\star$, as well as i, T_0, $a/R_\star$, the quadratic law limb-darkening coefficient u_1 and the systematics model parameters, the x-pixel and y-pixel shifts, FWHM and sky background, with the period P, the eccentricity e and u_2 fixed. We used uniform priors for all the fitted parameters. Our analytic transit light-curve model was generated with batman. We fitted our broadband light curve with a transit + systematics model using a Gaussian process (GP)[63,64], implemented through george[65], and a Markov chain Monte Carlo method, implemented through emcee[66]. For our Tiberius spectroscopic light curves, we held $a/R_\star$, i and T_0 fixed to the values determined from the broadband light-curve fits and applied a systematics correction from the broadband light-curve fit to aid in fitting the mirror-tilt event. We fitted the spectroscopic light curves using a GP with an exponential squared kernel for the same systematics detrending parameters detailed above. We used a Gaussian prior for $a/R_\star$ and uniform priors for all other fitted parameters.

Fitting pipeline 3: Aesop

We used transit light curves from the ExoTiC-JEDI [V1] stellar spectra and fit the broadband and spectroscopic light curves using the least-squares minimizer LMFIT[67]. We fitted each light curve with a two-component function consisting of a transit model (generated using batman) multiplied by a systematics model. Our systematics model included the x-pixel and y-pixel positions on the detector (x, y, xy, x^2 and y^2). To capture the amplitude

of the tilt event in our systematics model, we carried out piecewise linear regression on the out-of-transit baseline pre-tilt and post-tilt. We first fit the broadband light curve by fixing P and e and fitting for T_0, $a/R_\star$, i, $R_p/R_\star$, stellar baseline flux and systematic trends using wide uniform priors. For the spectroscopic light curves, we fixed T_0, $a/R_\star$ and i to the best-fit values from the broadband light curve and fit for $R_p/R_\star$. We held the nonlinear limb-darkening coefficients fixed.

Fitting pipeline 4: transitspectroscopy

We fit the broadband and spectroscopic light curves produced from the transitspectroscopy stellar spectra, running juliet[68] in parallel with the light-curve-fitting module of the transitspectroscopy pipeline[48] with dynamic nested sampling through dynesty[69] and analytical transit models computed using batman. We fit the broadband light curves for NRS1 and NRS2 individually, fixing P, e and ω and fitting for the impact parameter b, as well as T_0, $a/R_\star$, $R_p/R_\star$, extra jitter and the mean out-of-transit flux. We also fitted two linear regressors, a simple slope and a 'jump' (a regressor with zeros before the tilt event and ones after the tilt event), scaled to fit the data. We fitted the square-root-law limb-darkening coefficients using the Kipping sampling scheme. We fitted the spectroscopic light curves at the native resolution of the instrument, fixing T_0, $a/R_\star$ and b. We used the broadband light-curve systematics model for the spectroscopic light curve, with wide uniform priors for each wavelength bin, and set truncated normal priors for the square-root-law limb-darkening coefficients. We also fitted a jitter term added in quadrature to the error bars at each wavelength with a log-uniform prior between 10 and 1,000 ppm. We computed the mean of the limb-darkening coefficients by first computing the nonlinear coefficients from ATLAS models[70] and passing them through the SPAM algorithm[71]. We binned the data into 10-pixel-wavelength bins after fitting the native-resolution light curves.

Fitting pipeline 5: ExoTEP

We fitted the transit light curves from the Eureka! [V1] stellar spectra using the ExoTEP analysis framework[72,73,74,75]. ExoTEP uses batman to generate analytical light-curve models, adds an analytical instrument systematics model along with a photometric scatter parameter and fits for the best-fit parameters and their uncertainties using emcee. Before fitting, we cleaned the light curves by running ten iterations of 5σ clipping using a running median of window length 20 on the flux, x-pixel and y-pixel shifts and the 'ydriftwidth' data product from Eureka! Stage 3 (the average spatial PSF width at each integration). Our systematics model consisted of a linear trend in time with a 'jump' (constant offset) after the tilt event. The 'ydriftwidth' was used before the fit to locate the tilt event. We used a running median of 'ydriftwidth' to search for the largest offset and flagged every data point after the tilt event so that they would receive a constant 'jump' offset in our systematics model. We also removed the first point of the tilt event in our fits, as it was not captured by the 'jump' model. We fitted the broadband light curves, fitting for $R_p/R_\star$, photometric scatter, T_0, b, $a/R_\star$, the quadratic limb-darkening coefficients and the systematics model parameters (normalization constant, slope in time and constant 'jump' offset). We used uninformative flat priors on all the parameters. The orbital parameters were fixed to the best-fit broadband light curve values for the subsequent spectroscopic light-curve fits.

Fitting pipeline 6

We fitted transit light curves from the ExoTiC-JEDI [V1] stellar spectra using a custom lmfit light-curve-fitting code. The final systematic detrending model included a batman analytical transit model multiplied by a systematics model consisting of a linear stellar baseline term, a linear term for the x-pixel and y-pixel shifts and an exponential ramp function. The tilt event was accounted for by decorrelating the light curves with the y-pixel shifts, using a (1 + constant × y-shift) term with the constant fitted for in each light curve. For the broadband light-curve fits, we fixed P and fitted for T_0, i, $R_p/R_\star$, $a/R_\star$, x-pixel and y-pixel shifts and the exponential ramp amplitude and timescale. We fixed the nonlinear limb-darkening coefficients. For the spectroscopic light-curve fits, we fixed the values of T_0, i and $a/R_\star$ and the exponential ramp timescale to the broadband light-curve-fit values, and

fitted for $R_p/R_\star$, the x-pixel and y-pixel shifts and the ramp amplitude. Wide, uniform priors were used on all the fitting parameters in both the broadband and spectroscopic light-curve fits.

Fitting pipeline 7

We fitted transit light curves from the Eureka! [V2] stellar spectra, using PyLightcurve (ref. [75]) to generate the transit model with emcee as the sampler. We calculated the nonlinear four-parameter limb-darkening coefficients using ExoTHETyS (ref. [76]), which relies on PHOENIX 2012–2013 stellar models[77,78], and fixed these in our fits to the precomputed theoretical values. Our full transit + systematics model included a transit model multiplied by a second-order polynomial in the time domain. We accounted for the tilt event by subtracting the mean of the last 30 integrations of the pre-transit data from the mean of the first 30 integrations of the post-transit data, to account for the jump in flux, shifting the post-transit light curve upwards by the jump value. We fitted for the systematics (the parameters of the second-order polynomial), $R_p/R_\star$ and T_0. We used uniform priors for all the fitted parameters. We adopted the root mean square of the out-of-transit data as the error bars for the light-curve data points to account for the scatter in the data.

Fitting pipeline 8

We used the transit light curves generated from the ExoTiC-JEDI [V1] stellar spectra. We fit the broadband light curves with a batman transit model multiplied by a second-order systematics model as a function of x-pixel and y-pixel shifts. We fixed both of the quadratic limb-darkening coefficients for each wavelength bin. We fitted for $R_p/R_\star$, i, T_0 and $a/R_\star$, using wide uninformed priors, and ran our fits using emcee. For the spectroscopic light-curve fits, we fixed i and $a/R_\star$ to the broadband light-curve best-fit values and fitted for an extra error term added in quadrature.

Fitting pipeline 9

We used the transit light curves from the ExoTiC-JEDI [V1] stellar spectra. We fixed both of the quadratic limb-darkening coefficients and fitted the light curves with a batman transit model multiplied by a systematics model of a second-order function of x-pixel and y-pixel shifts. We fixed the best-fit broadband light-curve values for T_0, $a/R_\star$ and i for the spectroscopic light-curve fits and fitted for $R_p/R_\star$ using emcee for each 10-pixel bin, with the walkers initialized in a tight cluster around the best-fit solution from a Levenberg–Marquardt minimization. For both the broadband and spectroscopic light curves, we also fit for an extra per-point error term.

Fitting pipeline 10

We fitted the transit light curves from the ExoTiC-JEDI [V2] stellar spectra and performed our model fitting using automatic differentiation implemented with JAX (ref. [79]). We used a GP systematics model with a time-dependent Matérn ($v = 3/2$) kernel and a variable white-noise jitter term. The mean function consists of a linear trend in time plus a sigmoid function to account for the drop in measured flux that occurred mid-transit owing to the mirror-tilt event. For the transit model, we used the exoplanet package[80], making use of previously developed light-curve models[81,82]. For the GP systematics component, a generalization of the algorithm used by the celerite package[83] was adapted for JAX. We fixed both of the quadratic limb-darkening coefficients. For the initial broadband light-curve fit, both NRS1 and NRS2 were fitted simultaneously. All transit parameters were shared across both light curves, except for $R_p/R_\star$, which was allowed to vary for NRS1 and NRS2 independently. We fitted for T_0, the transit duration b and both $R_p/R_\star$ values. For the spectroscopic light-curve fits, all transit parameters were then fixed to the maximum-likelihood values determined from the broadband fit, except for $R_p/R_\star$, which was allowed to vary for each wavelength bin. Uncertainties for the transit model parameters, including $R_p/R_\star$, were assumed to be Gaussian and estimated using the Fisher information matrix at the location of the maximum-likelihood solution, which was computed exactly using the JAX automatic differentiation framework.

Fitting pipeline 11: Eureka!

We used transit light curves from the Eureka! [V2] time-series stellar spectra with the open-source Eureka! code to estimate the best-fit transit parameters and their uncertainties using a Markov chain Monte Carlo method fit to the data (implemented by emcee). A linear trend in time was used as a systematics model and a step function was used to account for the tilt event. We fixed $a/R_\star$, i, T_0 and the time of the tilt event to the best-fit values from the NRS1 broadband light curve, with the three integrations during the tilt event clipped from the light curve. We fitted for $R_p/R_\star$, both quadratic limb-darkening coefficients, the linear time trend and the magnitude of the step from the tilt event, with uniform priors for both the magnitude of the step and the limb-darkening coefficients.

Transmission spectral analysis

On the basis of the independent light-curve fits described above, we produced 11 transmission spectra from our NIRSpec G395H observations using several analyses and fitting methods. Extended Data Table 1 shows a breakdown of the different steps used in each fitting pipeline. In this work, three different 2D spectral image products were used, producing seven different 1D stellar spectra. Eleven different fitting pipelines using five different limb-darkening methods were then applied. Each of these fitting pipelines resulted in an independent analysis of the observations and 11 comparative transmission spectra. Extended Data Fig. 5 details comparative information for all 11 analyses to quantify their similarities and differences.

We computed the standard deviation of the 11 spectra in each wavelength bin and compared this to the mean uncertainty obtained in that bin. The average standard deviation in each bin across all fitting pipelines was 199 ppm, compared with an average uncertainty of 221 ppm (which ranged from 131 to 625 ppm across the bins). The computed standard deviation in each bin across all pipelines ranged from 85 to 1,040 ppm, with greater than 98% of the bins having a standard deviation lower than 500 ppm. We see an increase in scatter at longer wavelengths, with the structure of the scatter following closely with the measured stellar flux, for which throughput beyond 3.8 μm combines with decreasing stellar flux. The unweighted mean

uncertainty of all 11 transmission spectra follows a similar structure to the standard deviation, with the uncertainty increasing at longer wavelengths. The uncertainties from each fitting pipeline are consistent to within 3σ of each other, with the uncertainty per bin typically overestimated compared with the mean uncertainty across all reductions.

From all 11 transmission spectra, we computed a weighted-average transmission spectrum using the transit depth values from all reductions in each bin weighted by 1/variance ($1/\sigma^2$, in which σ is the uncertainty on the data point from each reduction). For this weighted-average transmission spectrum, the unweighted mean of the uncertainties in each bin was used to represent the error bar on each point. By using the weighted average of all 11 independently obtained transmission spectra, we therefore do not apply infinite weight to any one reduction in our interpretation of the atmosphere. Although the weighted average could be sensitive to any one spectrum with underestimated uncertainties, we find that our uncertainties are typically overestimated compared with the average. Similarly, we chose to use the mean rather than the median of the transmission spectral uncertainties, as this results in a more conservative estimate of the uncertainties in each bin. We find that all of the 11 transmission spectra are within 2.95σ of the weighted-average transmission spectrum without applying offsets.

We calculated normalized transmission spectrum residuals for each fitting pipeline by subtracting the weighted-average spectrum and dividing by the uncertainty in each bin. We generated histograms of the normalized transmission spectrum residuals and used the mean and standard deviation of the residuals to compute a normalized probability density function (PDF). We performed a Kolmogorov–Smirnov test on each of the normalized residuals and found that they are all approximately symmetric around their means, with normal distributions. This confirms that they are Gaussian in shape, with the null hypothesis that they are not Gaussian strongly rejected in the majority of cases (see Extended Data Fig. 5).

The PDFs of the residuals indicate three distinct clusters of computed spectra based on their deviations from the mean and their spreads. The first cluster is negatively offset by less than 200 ppm and corresponds to fitting pipelines that used extracted stellar spectra and that underwent further cleaning steps (for example, ExoTiC-JEDI [V3]). The second cluster is

positively offset from the mean by about 120 ppm and contains most of the transmission spectra produced. We see no obvious trends in this group to any specific reduction or fitting process. The final cluster is centred around the mean but has a broad distribution, suggesting a larger scatter both above and below the average transmission spectrum. This is probably the result of uncleaned outliers or marginal offsets between NRS1 and NRS2. These transmission spectra demonstrate that the 11 independent fitting pipelines are able to accurately reproduce the same transmission spectral feature structures, further highlighting the minimized impact of systematics on the time-series light curves. We suspect that the minor differences resulting from different reduction products and fitting pipelines are linked to the superbias and treatment of $1/f$ noise. We anticipate that the impacts of these will be improved with new superbias images, expected to be released soon by STScI, and with more detailed investigation into the impact of $1/f$ noise at the group level beyond the scope of this work.

Model comparison

To identify spectral absorption features, we compared the resulting weighted-average transmission spectrum of WASP-39b to several 1D RCTE atmosphere models from three independent model grids. Each forward model is computed on a set of common physical parameters (for example, metallicity, C/O ratio, internal temperature and heat redistribution), shown in Extended Data Table 2. Additionally, each model grid contains different prescriptions for treating certain physical effects (for example, scattering aerosols). Although each grid contains different opacity sources from varying line lists (see Extended Data Table 2), they each consider all of the main molecular and atomic species[84]. Each model transmission spectrum from the grids was binned to the same resolution as that of the observations to compute the χ^2 per data point, with a wavelength-independent transit depth offset as the free parameter. In general, the forward model grids fit the main features of the data but are unable to place statistically significant constraints on many of the atmospheric parameters, owing to both the finite nature of the forward model grid spacing[13] and the insensitivity of some of these parameters to the 3–5-μm transmission spectrum of WASP-39b (for example, >100 K differences in interior temperature provided nearly identical χ^2/N).

ATMO

We used the ATMO RCTE grid[85,86,87,88], which consists of model transmission spectra for different day–night energy redistribution factors, atmospheric metallicities, C/O ratios, haze factors and grey cloud factors with a range of line lists and pressure-broadening sources[88]. In total, there were 5,160 models. Within this grid, we find the best-fit model to have 3 times solar metallicity, with a C/O ratio of 0.35 and a grey cloud opacity 5 times the strength of H_2 Rayleigh scattering at 350 nm and a $\chi^2/N = 1.098$ for $N = 344$ data points and only fitting for an absolute altitude change in y.

PHOENIX

We calculated a grid of transmission spectra using the PHOENIX atmosphere model[89,90,91], varying the heat redistribution of the planet, atmospheric metallicity, C/O ratio, internal temperature, the presence of aerosols and the atmospheric chemistry (equilibrium or rainout). Opacities used include the BT2 H_2O line list[92], as well as HITRAN for 129 other main molecular absorbers[93] and Kurucz and Bell data for atomic species[94]. The HITRAN line lists available in this version of PHOENIX are often complete only at room temperature, which may be the cause of the apparent shift in the CO_2 spectral feature compared with the other grids that primarily use HITEMP and ExoMol lists. This shift is the cause of the difference in χ^2 between PHOENIX and the other model grids. In total, there were 1,116 models. Within this grid, the best-fit model has 10 times solar metallicity, a C/O ratio of 0.3, an internal temperature of 400 K, rainout chemistry and a cloud deck top at 0.3 mbar. The best-fit model has a $\chi^2/N = 1.203$ for $N = 344$ data points.

PICASO 3.0 and Virga

We used the open-source radiative–convective equilibrium model PICASO 3.0 (refs. [95,96]), which has its heritage in the Fortran-based EGP mode[97,98], to compute a grid of 1D pressure–temperature models for WASP-39b. The opacity sources included in PICASO 3.0 are listed in Extended Data Table 2.

Of the 29 molecular opacity sources included, the line lists of notable molecules used were: H_2O (ref. [99]), CO_2 (ref. [100]), CH_4 (ref. [101]) and CO (ref. [102]). The parameters varied in this grid of models include the interior temperature of the planet (T_{int}), atmospheric metallicity, C/O ratio and the dayside-to-nightside heat redistribution factor (see Extended Data Table 2), with correlated-*k* opacities[98,103]. In total, there were 192 cloud-free models. We include the effect of clouds in two ways. First, we post-processed the pressure–temperature profile using the cloud model Virga[95,104], which follows from previously developed methodologies[38], in which we included three condensable species (MnS, Na_2S and $MgSiO_3$). Virga requires a vertical mixing parameter, K_{zz} ($cm^2\ s^{-1}$), and a vertically constant sedimentation efficiency parameter, f_{sed}. In general, f_{sed} controls the vertical extent of the cloud opacity, with low values ($f_{sed} < 1$) creating large, vertically extended cloud decks with small particle sizes. In total, there were 3,840 cloudy models. The best fit from our grid with Virga-computed clouds has 3 times solar metallicity, solar C/O (0.458) and $f_{sed} = 0.6$, which results in a $\chi^2/N = 1.084$.

As well as the grid fit, we also use the PICASO framework to quantify the feature-detection significance. In this method, we are able to incorporate clouds on the fly using the fitting routine PyMultiNest[105]. We fit for each of the grid parameters using a nearest-neighbour technique and a radius scaling to account for the unknown reference pressure, giving five parameters in total. When fitting for clouds, we either fit for K_{zz} and f_{sed} in the Virga framework (seven parameters in total) or we fit for the cloud-top pressure corresponding to a grey cloud deck with infinite opacity (six parameters in total). These results are described in the following section.

Feature-detection significance

From the chemical equilibrium results of the single best-fit models, the molecules that could potentially contribute to the spectrum based on their abundances and 3–5-μm opacity sources are H_2 and He (via continuum) and CO, H_2O, H_2S, CO_2 and CH_4. More minor sources of opacity with VMR abundances <1 ppm are molecules such as OCS and NH_3. For example, removing H_2S, NH_3 and OCS from the single best-fit PICASO 3.0 model increases the chi-square value by less than 0.002. Therefore, we focus on computing the statistical significance of only H_2O, SO_2, CO_2, CH_4 and CO.

To quantify the statistical significance, we performed two different tests. First, we used a Gaussian residual fitting analysis, as used in other JTEC ERS analyses[23,29,31]. In this method, we subtracted the best-fit model without a specific opacity source from the weighted-average spectrum of WASP-39b, isolating the supposed spectral feature. We then fit a three-parameter or four-parameter Gaussian curve to the residual data using a nested sampling algorithm to calculate the Bayesian evidence[106]. For H_2O and CO, the extra transit depth offset parameter for the Gaussian fit was necessary to account for local mismatch of the fit to the continuum, whereas only a mean, standard deviation and scale parameter were required for a residual fit to the other molecules. We then compared this to the Bayesian evidence of a flat line to find the Bayes factor between a model that fits the spectral feature versus a model that excludes the spectral feature. These fits are shown in Extended Data Fig. 6.

Although the Gaussian residual fitting method is useful for quantifying the presence of potentially unknown spectral features, it cannot robustly determine the source of any given opacity. We therefore used the Bayesian fitting routine from PyMultiNest in the PICASO 3.0 framework to refit the grid parameters, while excluding the opacity contribution from the species in question. Then, we compared the significance of the molecule through a Bayes factor analysis[107]. Those values are shown in Extended Data Table 3.

We find significant evidence (>3σ) for H_2O, CO_2 and SO_2. In general, the two methods only agree well for molecules whose contribution has a Gaussian shape (that is, SO_2 and CO_2). For example, for CO_2, we find decisive 28.5σ and 26.9σ detections for the Bayes factor and Gaussian analysis, respectively. Similarly, for H_2O, we find 21.5σ and 16.5σ detections, respectively. The evidence for SO_2 is less substantial, but both methods give significant detections of 4.8σ and 3.5σ, respectively. Although the Gaussian fitting method found a broad 1-μm-wide residual in the region of CO (that is, >4.5 μm), its shape was unlike that seen with the PRISM data[31]. CO remained undetected with the Bayesian fitting analysis and therefore we are unable to robustly confirm evidence of CO. Similarly, no evidence for CH_4 was found[23]. Gaussian residual fitting in the region of CH_4 absorption only found a very broad inverse Gaussian and so is not included in Extended Data Table 3.

SO_2 absorption

We performed an injection test with the PICASO best-fit model in the PyMultiNest fitting framework to determine the abundance of SO_2 required to match the observations. We add SO_2 opacity using the ExoMol line list[108], without rerunning the RCTE model to self-consistently compute a new climate profile. Fitting for the cloud deck dynamically, without SO_2, produces a single best estimate of 10 times solar metallicity, sub-solar C/O (0.229), resulting in a marginally worse $\chi^2/N = 1.11$. With SO_2, the single best fit tends back to 3 times solar metallicity, solar C/O. This suggests that cloud treatment and the exclusion of spectrally active molecules have an effect on the resultant physical interpretation of bulk atmospheric parameters. Ultimately, if we fit for SO_2 in our PyMultiNest framework with the Virga cloud treatment, we obtain 3 times solar metallicity, solar C/O, $\log SO_2 = -5.6 \pm 0.1$ ($SO_2 = 2.5 \pm 0.65$ ppm) and $\chi^2/N = 1.02$, which is our single best-fit model (shown in Fig. 4). For context, an atmospheric metallicity of 3–10 times solar would provide a thermochemical equilibrium abundance of 72–240 ppm H_2S, the presumed source for photochemically produced SO_2 (ref. [36]).

To confirm the plausibility of SO_2 absorption to explain the 4.1-µm spectral feature, we also computed models with prescribed, vertically uniform SO_2 VMRs of 0, 1, 5 and 10 ppm using the structure from the best-fit PHOENIX model (10 times solar metallicity, C/O = 0.3). We calculated ad hoc spectra using the gCMCRT radiative transfer code[109] with the ExoMol SO_2 line list[108] (see Extended Data Fig. 7). Linearly interpolating the models with respect to the SO_2 abundance and performing a Levenberg–Marquardt regression gave a best-fit value of 4.6 ± 0.67 ppm. Inserting this abundance of SO_2 into the best-fit PHOENIX model improves the χ^2/N from 1.2 to 1.08.

Future atmospheric retrievals can provide a more statistically robust measurement for the SO_2 abundance and add extra information from the similar absorption seen in the PRISM transmission spectrum[29,31].

4.56-µm feature

A 0.08-µm-wide bump in transit depth centred at 4.56 µm is not fit by any of the model grids. This feature, picked up by the resolution of G395H, is not clearly seen in other ERS observations of WASP-39b. Following the same Gaussian residual fitting procedure as described above, we found a feature significance of 3.3σ (see Extended Data Fig. 6). To identify possible opacity sources in the atmosphere of WASP-39b that might be the cause of this absorption, we compared the feature with CH_4 (ref. 110), C_2H_2 (ref. 111), C_2H_4 (ref. 112), C_2H_6 (ref. 113), CO (ref. 114), CO_2 (ref. 100), CS_2 (ref. 113), CN (ref. 115), HCN (ref. 116), HCl (ref. 113), H_2S (ref. 117), HF (ref. 118), H_3^+ (ref. 119), LiCl (ref. 115), NH_3 (ref. 120), NO (ref. 114), NO_2 (ref. 113), N_2O (ref. 114), N_2 (ref. 121), NaCl (ref. 122), OCS (ref. 113), PH_3 (ref. 123), PN (ref. 124), PO (ref. 125), SH (ref. 126), SiS (ref. 127), SiH_4 (ref. 128), SiO (ref. 129), the X–X state of SO (ref. 130), SO_2 (ref. 108), SO_3 (ref. 108) and isotopologues of H_2O, CH_4, CO_2 and CO, but did not find a convincing candidate that showed opacity at the correct wavelength or the correct width. The narrowness of the feature suggests that it could be a very distinct Q-branch, in which the rotational quantum number in the ground state is the

same as the rotational quantum number in the excited state. However, of the molecules we explored, there were no candidates with a distinct Q-branch at this wavelength whose P-branch and R-branch did not obstruct the neighbouring CO_2 and continuum-like CO + H_2O opacity.

We also note that many of these species lack high-temperature line-list data, making it difficult to definitively rule out such species. For example, OCS, SO and CS_2 are available in HITRAN2020 (ref. [113]) but not in ExoMol[131]. Furthermore, if photochemistry is important for WASP-39b, as indicated by the presence of SO_2, then there may be many species out of equilibrium that may contribute to the transit spectrum, some of which do not have high-temperature opacity data at present (such as OCS, NH_2 or HSO). Future observations over this wavelength region of this and other planets may confirm or refute the presence of this unknown absorber.

Data availability

The data used in this paper are associated with JWST programme ERS 1366 (observation #4) and are available from the Mikulski Archive for Space Telescopes (MAST; https://mast.stsci.edu). Science data processing version (SDP_VER) 2022_2a generated the uncalibrated data that we downloaded from MAST. We used JWST Calibration Pipeline software version (CAL_VER) 1.5.3 with modifications described in the text. We used calibration reference data from context (CRDS_CTX) 0916, except as noted in the text. All the data and models presented in this publication can be found at https://doi.org/10.5281/zenodo.7185300.

Code availability

The codes used in this publication to extract, reduce and analyse the data are as follows; STScI JWST Calibration Pipeline[44] (https://github.com/spacetelescope/jwst), Eureka![50] (https://eurekadocs.readthedocs.io/en/latest/), ExoTiC-JEDI[45] (https://github.com/Exo-TiC/ExoTiC-JEDI), juliet[68] (https://juliet.readthedocs.io/en/latest/), Tiberius[15,46,47],

transitspectroscopy[48] (https://github.com/nespinoza/transitspectroscopy). Furthermore, these made use of batman[62] (http://lkreidberg.github.io/batman/docs/html/index.html), celerite[83] (https://celerite.readthedocs.io/en/stable/), chromatic (https://zkbt.github.io/chromatic/), dynesty[69] (https://dynesty.readthedocs.io/en/stable/index.html), emcee[66] (https://emcee.readthedocs.io/en/stable/), exoplanet[80] (https://docs.exoplanet.codes/en/latest/), ExoTEP[72,73,74], ExoTHETyS[76] (https://github.com/ucl-exoplanets/ExoTETHyS), ExoTiC-ISM[54] (https://github.com/Exo-TiC/ExoTiC-ISM), ExoTiC-LD[55] (https://exotic-ld.readthedocs.io/en/latest/), george[65] (https://george.readthedocs.io/en/latest/), JAX[79] (https://jax.readthedocs.io/en/latest/), LMFIT[67] (https://lmfit.github.io/lmfit-py/), PyLightcurve[75] (https://github.com/ucl-exoplanets/pylightcurve), PyMC3 (ref. [132]) (https://docs.pymc.io/en/v3/index.html) and Starry[81] (https://starry.readthedocs.io/en/latest/), each of which use the standard Python libraries astropy[133,134], matplotlib[135], numpy[136], pandas[137], scipy[61] and xarray[138]. The atmospheric models used to fit the data can be found at ATMO[85,86,87,88], PHOENIX[89,90,91], PICASO[95,96] (https://natashabatalha.github.io/picaso/), Virga[95,104] (https://natashabatalha.github.io/virga/) and gCMCRT[109] (https://github.com/ELeeAstro/gCMCRT).

References

1. Öberg, K. I., Murray-Clay, R. & Bergin, E. A. The effects of snowlines on C/O in planetary atmospheres. *Astrophys. J. Lett.* **743**, L16 (2011).

2. Mordasini, C., van Boekel, R., Mollière, P., Henning, T. & Benneke, B. The imprint of exoplanet formation history on observable present-day spectra of hot Jupiters. *Astrophys. J.* **832**, 41 (2016).

3. Sing, D. K. et al. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature* **529**, 59–62

(2016).

4. Wakeford, H. R. et al. The complete transmission spectrum of WASP-39b with a precise water constraint. *Astron. J.* **155**, 29 (2018).

5. Alam, M. K. et al. The Hubble Space Telescope PanCET program: an optical to infrared transmission spectrum of HAT-P-32Ab. *Astron. J* **160**, 51 (2020).

6. Birkby, J. L. Exoplanet atmospheres at high spectral resolution. Preprint at https://arxiv.org/abs/1806.04617 (2018).

7. Line, M. R. et al. A solar C/O and sub-solar metallicity in a hot Jupiter atmosphere. *Nature* **598**, 580–584 (2021).

8. Pelletier, S. et al. Where is the water? Jupiter-like C/H ratio but strong H_2O depletion found on τ Boötis b using SPIRou. *Astron. J* **162**, 73 (2021).

9. Faedi, F. et al. WASP-39b: a highly inflated Saturn-mass planet orbiting a late G-type star. *Astron. Astrophys.* **531**, A40 (2011).

10. Batalha, N. E. & Line, M. R. Information content analysis for selection of optimal JWST observing modes for transiting exoplanet atmospheres. *Astron. J* **153**, 151 (2017).

11. Jakobsen, P. et al. The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. I. Overview of the instrument and its capabilities. *Astron. Astrophys.* **661**, A80 (2022).

12. Birkmann, S. M. et al. The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. IV. Capabilities and predicted performance for exoplanet characterization. *Astron. Astrophys.* **661**, A83 (2022).

13. Fischer, P. D. et al. *HST* hot-Jupiter transmission spectral survey: clear skies for cool Saturn WASP-39b. *Astrophys. J.* **827**, 19 (2016).

14. Nikolov, N. et al. VLT FORS2 comparative transmission spectroscopy: detection of Na in the atmosphere of WASP-39b from the ground. *Astrophys. J.* **832**, 191 (2016).

15. Kirk, J. et al. LRG-BEASTS: transmission spectroscopy and retrieval analysis of the highly inflated Saturn-mass planet WASP-39b. *Astron. J.* **158**, 144 (2019).

16. Barstow, J. K., Aigrain, S., Irwin, P. G. & Sing, D. K. A consistent retrieval analysis of 10 hot Jupiters observed in transmission. *Astrophys. J.* **834**, 50 (2017).

17. Pinhas, A., Madhusudhan, N., Gandhi, S. & MacDonald, R. H_2O abundances and cloud properties in ten hot giant exoplanets. *Mon. Not. R. Astron. Soc.* **482**, 1485–1498 (2019).

18. Tsiaras, A. et al. A population study of gaseous exoplanets. *Astron. J* **155**, 4 (2018).

19. Welbanks, L. et al. Mass–metallicity trends in transiting exoplanets from atmospheric abundances of H_2O, Na, and K. *Astrophys. J. Lett.* **887**, L20 (2019).

20. Kawashima, Y. & Min, M. Implementation of disequilibrium chemistry to spectral retrieval code ARCiS and application to 16 exoplanet transmission spectra. Indication of disequilibrium chemistry for HD 209458b and WASP-39b. *Astron. Astrophys.* **656**, A90 (2021).

21. Shibata, S., Helled, R. & Ikoma, M. The origin of the high metallicity of close-in giant exoplanets. Combined effects of resonant and aerodynamic shepherding. *Astron. Astrophys.* **633**, A33 (2020).

22. Helled, R. & Morbidelli, A. in *ExoFrontiers: Big Questions in Exoplanetary Science* (ed. Madhusudhan, N.) (IOP Publishing, 2021).

23. Ahrer, E.-M. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRCam. *Nature* https://doi.org/10.1038/s41586-022-05590-4 (2023).

24. Polanski, A. S., Crossfield, I. J., Howard, A. W., Isaacson, H. & Rice, M. Chemical abundances for 25 JWST exoplanet host stars with KeckSpec. *Res. Notes AAS* **6**, 155 (2022).

25. Mullally, S. E., Rodriguez, D. R., Stevenson, K. B. & Wakeford, H. R. The Exo.MAST table for JWST exoplanet atmosphere observability. *Res. Notes AAS* **3**, 193 (2019).

26. Stevenson, K. B. et al. Transiting exoplanet studies and community targets for JWST's Early Release Science Program. *Publ. Astron. Soc. Pac.* **128**, 094401 (2016).

27. Bean, J. L. et al. The Transiting Exoplanet Community Early Release Science Program for JWST. *Publ. Astron. Soc. Pac.* **130**, 114402 (2018).

28. Rigby, J. et al. The science performance of JWST as characterized in commissioning. Preprint at https://arxiv.org/abs/2207.05632 (2023).

29. JWST Transiting Exoplanet Community Early Release Science Team. Identification of carbon dioxide in an exoplanet atmosphere. *Nature* https://doi.org/10.1038/s41586-022-05269-w (2022).

30. Lodders, K. & Fegley, B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars: I. Carbon, nitrogen, and oxygen. *Icarus* **155**, 393–424 (2002).

31. Rustamkulov, R. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM. *Nature* https://doi.org/10.1038/s41586-022-05677-y (2023).

32. Zahnle, K. et al. Atmospheric sulfur photochemistry on hot Jupiters. *Astrophys. J. Lett.* **701**, L20 (2009).

33. Hobbs, R. et al. Sulfur chemistry in the atmospheres of warm and hot Jupiters. *Mon. Not. R. Astron. Soc.* **506**, 3186–3204 (2021).

34. Tsai, S.-M. et al. A comparative study of atmospheric chemistry with VULCAN. *Astrophys. J.* **923**, 264 (2021).

35. Polman, J., Waters, L.B.F.M., Min, M., Miguel, Y. & Khorshid, N. H2S and SO2 detectability in hot Jupiters. Sulphur species as indicators of metallicity and C/O ratio. *Astron. Astrophys.* https://doi.org/10.1051/0004-6361/202244647 (in the press).

36. Moses, J. I., Madhusudhan, N., Visscher, C. & Freedman, R. S. Chemical consequences of the C/O ratio on hot Jupiters: examples from WASP-12b, CoRoT-2b, XO-1b, and HD 189733b. *Astrophys. J.* **763**, 25 (2012).

37. Moses, J. I. et al. Disequilibrium carbon, oxygen, and nitrogen chemistry in the atmospheres of HD 189733b and HD 209458b. *Astrophys. J.* **737**, 15 (2011).

38. Ackerman, A. S. & Marley, M. S. Precipitating condensation clouds in substellar atmospheres. *Astrophys. J.* **556**, 872 (2001).

39. Mousis, O., Aguichine, A., Helled, R., Irwin, P. G. J. & Lunine, J. I. The role of ice lines in the formation of Uranus and Neptune. *Philos. Trans. R. Soc. A* **378**, 20200107 (2020).

40. Wong, M. H., Mahaffy, P. R., Atreya, S. K., Niemann, H. B. & Owen, T. C. Updated Galileo probe mass spectrometer measurements of carbon, oxygen, nitrogen, and sulfur on Jupiter. *Icarus* **171**, 153–170 (2004).

41. Fletcher, L. N., Orton, G. S., Teanby, N. A., Irwin, P. G. J. & Bjoraker, G. L. Methane and its isotopologues on Saturn from Cassini/CIRS observations. *Icarus* **199**, 351–367 (2009).

42. Pollack, J. B. et al. Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).

43. Thorngren, D. & Fortney, J. J. Connecting giant planet atmosphere and interior modeling: constraints on atmospheric metal enrichment.

Astrophys. J. Lett. **874**, L31 (2019).

44. Bushouse, H. et al. JWST Calibration Pipeline (1.6.2). *Zenodo* https://doi.org/10.5281/zenodo.7041998 (2022).

45. Alderson, L., Grant, D., Wakeford, H. Exo-TiC/ExoTiC-JEDI: v0.1-beta-release. *Zenodo* https://doi.org/10.5281/zenodo.7185855 (2022).

46. Kirk, J. et al. LRG-BEASTS III: ground-based transmission spectrum of the gas giant orbiting the cool dwarf WASP-80. *Mon. Not. R. Astron. Soc.* **474**, 876 (2018).

47. Kirk, J. et al. ACCESS and LRG-BEASTS: a precise new optical transmission spectrum of the ultrahot Jupiter WASP-103b. *Astronom. J.* **162**, 34 (2021).

48. Espinoza, N. TransitSpectroscopy (0.3.11). *Zenodo* https://doi.org/10.5281/zenodo.6960924 (2022).

49. Marsh, T. R. The extraction of highly distorted spectra. *Publ. Astron. Soc. Pac.* **101**, 1032 (1989).

50. Bell, T. J. et al. Eureka!: An end-to-end pipeline for JWST time-series observations. *J. Open Source Softw.* **7**, 4503 (2022).

51. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for $-5.0 \leq \log[M/H] \leq +1$, 2000 K $\leq$ $T_{eff} \leq$ 50000 K at several surface gravities. *Astron. Astrophys.* **363**, 1081–1190 (2000).

52. Magic, Z., Chiavassa, A., Collet, R. & Asplund, M. The STAGGER-grid: a grid of 3D stellar atmosphere models. IV. Limb darkening coefficients. *Astron. Astrophys.* **573**, A90 (2015).

53. Space Telescope Science Institute. JWST User Documentation (JDox) website. https://jwst-docs.stsci.edu/ (2016).

54. Laginja, I. & Wakeford, H. ExoTiC-ISM: a Python package for marginalised exoplanet transit parameters across a grid of systematic instrument models. *J. Open Source Softw.* **5**, 2281 (2020).

55. Wakeford, H. & Grant, D. Exo-TiC/ExoTiC-LD: ExoTiC-LD v2.1 *Zenodo* https://doi.org/10.5281/zenodo.6809899 (2022).

56. Sing, D. K. Stellar limb-darkening coefficients for CoRot and Kepler. *Astron. Astrophys.* **510**, A21 (2010).

57. Kipping, D. M. Efficient, uninformative sampling of limb-darkening coefficients for two-parameter laws. *Mon. Not. R. Astron. Soc.* **435**, 2152–2160 (2013).

58. Espinoza, N. & Jordán, A. Limb darkening and exoplanets – II. Choosing the best law for optimal retrieval of transit parameters. *Mon. Not. R. Astron. Soc.* **457**, 3573–3581 (2016).

59. Allan, D. W. Statistics of atomic frequency standards. *Proc. IEEE* **54**, 221–230 (1966).

60. Pont, F., Zucker, S. & Queloz, D. The effect of red noise on planetary transit detection. *Mon. Not. R. Astron. Soc.* **373**, 231–242 (2006).

61. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

62. Kreidberg, L. batman: BAsic Transit Model cAlculatioN in Python. *Publ. Astron. Soc. Pac.* **127**, 1161 (2015).

63. Gibson, N. P. et al. A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. *Mon. Not. R. Astron. Soc.* **419**, 2683–2694 (2012).

64. Gibson, N. P. Reliable inference of exoplanet light-curve parameters using deterministic and stochastic systematics models. *Mon. Not. R. Astron. Soc.* **445**, 3401–3414 (2014).

65. Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W. & O'Neil, M. Fast direct methods for Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 252–265 (2015).

66. Foreman-Mackey, D. et al. emcee v3: a Python ensemble sampling toolkit for affine-invariant MCMC. *J. Open Source Softw.* **4**, 1864 (2019).

67. Newville, M. et al. LMFIT: non-linear least-square minimization and curve-fitting for Python. *Zenodo* https://doi.org/10.5281/zenodo.11813 (2014).

68. Espinoza, N., Kossakowski, D. & Brahm, R. juliet: a versatile modelling tool for transiting and non-transiting exoplanetary systems. *Mon. Not. R. Astron. Soc.* **490**, 2262–2283 (2019).

69. Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158 (2020).

70. Kurucz, R. L. Model atmospheres for g, f, a, b, and o stars. *Astrophys. J. Suppl. Ser.* **40**, 1–340 (1979).

71. Howarth, I. D. On stellar limb darkening and exoplanetary transits. *Mon. Not. R. Astron. Soc.* **418**, 1165–1175 (2011).

72. Benneke, B. et al. Spitzer observations confirm and rescue the habitable-zone super-Earth K2-18b for future characterization. *Astrophys. J.* **834**, 187 (2017).

73. Benneke, B. et al. A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds. *Nat. Astron.* **3**, 813–821 (2019).

74. Benneke, B. et al. Water vapor and clouds on the habitable-zone sub-Neptune exoplanet K2-18b. *Astrophys. J. Lett.* **887**, L14 (2019).

75. Tsiaras, A. et al. Detection of an atmosphere around the super-Earth 55 Cancri e. *Astrophys. J.* **820**, 99 (2016).

76. Morello, G. et al. ExoTETHyS: tools for exoplanetary transits around host stars. *J. Open Source Softw.* **5**, 1834 (2020).

77. Claret, A., Hauschildt, P. H. & Witte, S. New limb-darkening coefficients for PHOENIX/1D model atmospheres. I. Calculations for 1500 K ≤ T_{eff} ≤ 4800 K *Kepler*, CoRot, *Spitzer*, *uvby*, *UBVRIJHK*, Sloan, and 2MASS photometric systems. *Astron. Astrophys.* **546**, A14 (2012).

78. Claret, A., Hauschildt, P. H. & Witte, S. New limb-darkening coefficients for PHOENIX/1D model atmospheres. II. Calculations for 5000 K ≤ T_{eff} ≤ 10 000 K *Kepler*, CoRot, *Spitzer*, *uvby*, *UBVRIJHK*, Sloan, and 2MASS photometric systems. *Astron. Astrophys* **552**, A16 (2013).

79. Bradbury, J. et al. JAX: Autograd and XLA. Astrophysics Source Code Library. https://ascl.net/2111.002 (2021).

80. Foreman-Mackey, D. et al. exoplanet-dev/exoplanet: exoplanet v0.5.0. *Zenodo* https://doi.org/10.5281/zenodo.4737444 (2020).

81. Luger, R. et al. STARRY: analytic occultation light curves. *Astron. J* **157**, 64 (2019).

82. Agol, E., Luger, R. & Foreman-Mackey, D. Analytic planetary transit light curves and derivatives for stars with polynomial limb darkening. *Astron. J* **159**, 123 (2020).

83. Foreman-Mackey, D., Agol, E., Ambikasaran, S. & Angus, R. Fast and scalable Gaussian process modeling with applications to astronomical time series. *Astron. J* **154**, 220 (2017).

84. Sharp, C. & Burrows, A. Atomic and molecular opacities for brown dwarf and giant planet atmospheres. *Astrophys. J. Suppl. Ser.* **168**, 140 (2007).

85. Tremblin, P. et al. Fingering convection and cloudless models for cool brown dwarf atmospheres. *Astrophys. J. Lett.* **804**, L17 (2015).

86. Drummond, B. et al. The effects of consistent chemical kinetics calculations on the pressure-temperature profiles and emission spectra of hot Jupiters. *Astron. Astrophys.* **594**, A69 (2016).

87. Goyal, J. M. et al. A library of ATMO forward model transmission spectra for hot Jupiter exoplanets. *Mon. Not. R. Astron. Soc.* **474**, 5158 (2018).

88. Goyal, J. M. et al. A library of self-consistent simulated exoplanet atmospheres. *Mon. Not. R. Astron. Soc.* **498**, 4680 (2020).

89. Hauschildt, P. H., Allard, F. & Baron, E. The NextGen model atmosphere grid for $3000 \leq T_{\mathrm{eff}} \leq 10{,}000$ K. *Astrophys. J.* **512**, 377 (1999).

90. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885–895 (2001).

91. Lothringer, J. D. & Barman, T. The PHOENIX exoplanet retrieval algorithm and using H^- opacity as a probe in ultrahot Jupiters. *Astron. J* **159**, 289 (2020).

92. Barber, R. J., Tennyson, J., Harris, G. J. & Tolchenov, R. N. A high-accuracy computed water line list. *Mon. Not. R. Astron. Soc.* **368**, 1087–1094 (2006).

93. Rothman, L. S. et al. The HITRAN 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 533–572 (2009).

94. Kurucz, R. & Bell, B. Atomic line data. CD-ROM no. 23 (Smithsonian Astrophysical Observatory, 1995).

95. Batalha, N. E., Marley, M. S., Lewis, N. K. & Fortney, J. J. Exoplanet reflected-light spectroscopy with PICASO. *Astrophys. J.* **878**, 70 (2019).

96. Mukherjee, S., Batalha, N. E., Fortney, J. J. & Marley, M. S. PICASO 3.0: a one-dimensional climate model for giant planets and brown dwarfs. *Astrophys. J.* **942**, 71 (2023).

97. Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D. & Freedman, R. Comparative planetary atmospheres: models of TrES-1 and HD 209458b. *Astrophys. J.* **627**, L69–L72 (2005).

98. Marley, M. S. et al. The Sonora brown dwarf atmosphere and evolution models. I. Model description and application to cloudless atmospheres in rainout chemical equilibrium. *Astrophys. J.* **920**, 85 (2021).

99. Polyansky, O. L. et al. ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *Mon. Not. R. Astron. Soc.* **480**, 2597–2608 (2018).

100. Huang, X., Gamache, R. R., Freedman, R. S., Schwenke, D. W. & Lee, T. J. Reliable infrared line lists for 13 CO_2 isotopologues up to E'=18,000 cm^{-1} and 1500 K, with line shape parameters. *J. Quant. Spectrosc. Radiat. Transfer* **147**, 134–144 (2014).

101. Yurchenko, S. N., Amundsen, D. S., Tennyson, J. & Waldmann, I. P. A hybrid line list for CH_4 and hot methane continuum. *Astron. Astrophys.* **605**, A95 (2017).

102. Li, G. et al. Rovibrational line lists for nine isotopologues of the CO molecule in the $X^1\Sigma^+$ ground electronic state. *Astrophys. J. Suppl. Ser.* **216**, 15 (2015).

103. Lupu, R., Freedman, R., Gharib-Nezhad, E., Visscher, C. & Molliere, P. Correlated k coefficients for H2-He atmospheres; 196 spectral windows and 1460 pressure-temperature points. *Zenodo* https://doi.org/10.5281/zenodo.5590989 (2021).

104. Rooney, C. M., Batalha, N. E., Gao, P. & Marley, M. S. A new sedimentation model for greater cloud diversity in giant exoplanets and brown dwarfs. *Astrophys. J.* **925**, 33 (2022).

105. Buchner, J. PyMultiNest: Python interface for MultiNest. Astrophysics Source Code Library. https://www.ascl.net/1606.005 (2016).

106. Skilling, J. Nested sampling. *AIP Conf. Proc.* **735**, 395–405 (2004).

107. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104 (2008).

108. Underwood, D. S. et al. ExoMol molecular line lists – XIV. The rotation–vibration spectrum of hot SO_2. *Mon. Not. R. Astron. Soc.* **459**, 3890–3899 (2016).

109. Lee, E. K. et al. 3D radiative transfer for exoplanet atmospheres. gCMCRT: a GPU-accelerated MCRT code. *Astrophys. J.* **929**, 180 (2022).

110. Hargreaves, R. J. et al. An accurate, extensive, and practical line list of methane for the HITEMP database. *Astrophys. J. Suppl. Ser.* **247**, 55 (2020).

111. Chubb, K. L., Tennyson, J. & Yurchenko, S. N. ExoMol molecular line lists – XXXVII. Spectra of acetylene. *Mon. Not. R. Astron. Soc.* **493**, 1531–1545 (2020).

112. Mant, B. P., Yachmenev, A., Tennyson, J. & Yurchenko, S. N. ExoMol molecular line lists – XXVII. Spectra of C_2H_4. *Mon. Not. R. Astron. Soc.* **478**, 3220–3232 (2018).

113. Gordon, I. E. et al. The HITRAN2020 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer* **277**, 107949 (2022).

114. Rothman, L. S. et al. HITEMP, the high-temperature molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transfer* **111**, 2139–2150 (2010).

115. Bernath, P. F. MoLLIST: molecular line lists, intensities and spectra. *J. Quant. Spectrosc. Radiat. Transfer* **240**, 106687 (2020).

116. Harris, G. J., Tennyson, J., Kaminsky, B. M., Pavlenko, Y. V. & Jones, H. R. A. Improved HCN/HNC linelist, model atmospheres and synthetic spectra for WZ Cas. *Mon. Not. R. Astron. Soc.* **367**, 400–406 (2006).

117. Azzam, A. A. A., Tennyson, J., Yurchenko, S. N. & Naumenko, O. V. ExoMol molecular line lists – XVI. The rotation–vibration spectrum of hot H_2S. *Mon. Not. R. Astron. Soc.* **460**, 4063–4074 (2016).

118. Coxon, J. A. & Hajigeorgiou, P. G. Improved direct potential fit analyses for the ground electronic states of the hydrogen halides: HF/DF/TF, HCl/DCl/TCl, HBr/DBr/TBr and HI/DI/TI. *J. Quant. Spectrosc. Radiat. Transfer* **151**, 133–154 (2015).

119. Mizus, I. I. et al. ExoMol molecular line lists – XX. A comprehensive line list for H_3^+. *Mon. Not. R. Astron. Soc.* **468**, 1717–1725 (2017).

120. Coles, P. A., Yurchenko, S. N. & Tennyson, J. ExoMol molecular line lists – XXXV. A rotation–vibration line list for hot ammonia. *Mon. Not. R. Astron. Soc.* **490**, 4638–4647 (2019).

121. Western, C. M. et al. The spectrum of N_2 from 4,500 to 15,700 cm^{-1} revisited with PGOPHER. *J. Quant. Spectrosc. Radiat. Transfer* **219**, 127–141 (2018).

122. Barton, E. J. et al. ExoMol molecular line lists V: the ro-vibrational spectra of NaCl and KCl. *Mon. Not. R. Astron. Soc.* **442**, 1821–1829 (2014).

123. Sousa-Silva, C., Al-Refaie, A. F., Tennyson, J. & Yurchenko, S. N. ExoMol line lists – VII. The rotation–vibration spectrum of phosphine up to 1500 K. *Mon. Not. R. Astron. Soc.* **446**, 2337–2347 (2015).

124. Yorke, L., Yurchenko, S. N., Lodi, L. & Tennyson, J. Exomol molecular line lists – VI. A high temperature line list for phosphorus nitride. *Mon. Not. R. Astron. Soc.* **445**, 1383–1391 (2014).

125. Prajapat, L. et al. ExoMol molecular line lists – XXIII. Spectra of PO and PS. *Mon. Not. R. Astron. Soc.* **472**, 3648–3658 (2017).

126. Gorman, M. N., Yurchenko, S. N. & Tennyson, J. ExoMol molecular line lists XXXVI: $X\,^2\Pi - X\,^2\Pi$ and $A\,^2\Sigma^+ - X\,^2\Pi$ transitions of SH. *Mon. Not. R. Astron. Soc.* **490**, 1652–1665 (2019). (2019).

127. Upadhyay, A., Conway, E. K., Tennyson, J. & Yurchenko, S. N. ExoMol line lists XXV: a hot line list for silicon sulphide, SiS. *Mon. Not. R. Astron. Soc.* **477**, 1520–1527 (2018).

128. Owens, A., Yachmenev, A., Thiel, W., Tennyson, J. & Yurchenko, S. N. ExoMol line lists – XXII. The rotation–vibration spectrum of silane up to 1200 K. *Mon. Not. R. Astron. Soc.* **471**, 5025–5032 (2017).

129. Barton, E. J., Yurchenko, S. N. & Tennyson, J. ExoMol line lists – II. The ro-vibrational spectrum of SiO. *Mon. Not. R. Astron. Soc.* **434**, 1469–1475 (2013).

130. Brady, R. P., Yurchenko, S. N., Kim, G. S., Somogyi, W. & Tennyson, J. An *ab initio* study of the rovibronic spectrum of sulphur monoxide (SO): diabatic vs. adiabatic representation. *Phys. Chem. Chem. Phys.* **24**, 24076–24088 (2022).

131. Tennyson, J. et al. The 2020 release of the ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J. Quant. Spectrosc. Radiat. Transfer* **255**, 107228 (2020).

132. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).

133. The Astropy Collaboration et al.Astropy: a community Python package for astronomy. *Astron. Astrophys.* **558**, A33 (2013).

134. The Astropy Collaboration et al.The Astropy Project: building an open-science project and status of the v2.0 core package. *Astron. J* **156**, 123 (2018).

135. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

136. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

137. McKinney, W. pandas: a foundational Python library for data analysis and statistics. In *PyHPC 2011: Python for High Performance and Scientific Computing* Vol. 14, 1–9 (2011).

138. Hoyer, S. & Hamman, J. xarray: N-D labeled arrays and datasets in Python. *J. Open Res. Softw.* **5**, 10 (2017).

139. Espinoza, N. et al. Spectroscopic time-series performance of JWST/NIRSpec from commissioning observations. Preprint at https://arxiv.org/abs/2211.01459 (2022).

140. Caffau, E., Ludwig, H. G., Steffen, M., Freytag, B. & Bonifacio, P. Solar chemical abundances determined with a CO5BOLD 3D model atmosphere. *Sol. Phys.* **258**, 255–269 (2011).

141. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. Preprint at https://arxiv.org/abs/0909.0948 (2022).

142. Lodders, K., Palme, H. & Gail, H.-P. Abundances of the elements in the solar system. Preprint at https://arxiv.org/abs/0901.1149 (2022).

Acknowledgements

This work is based on observations made with the NASA/ESA/CSA JWST. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for the JWST. These observations are associated with programme JWST-ERS-01366. Support for programme JWST-ERS-01366 was provided by NASA through a grant from the Space Telescope Science

Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127. L.A. acknowledges funding from STFC grant ST/W507337/1 and from the University of Bristol School of Physics PhD Scholarship Fund.

Author information

Authors and Affiliations

1. School of Physics, HH Wills Physics Laboratory, University of Bristol, Bristol, UK

   Lili Alderson, Hannah R. Wakeford & David Grant

2. Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA

   Munazza K. Alam & Nicole L. Wallack

3. NASA Ames Research Center, Moffett Field, CA, USA

   Natasha E. Batalha

4. Department of Physics, Utah Valley University, Orem, UT, USA

   Joshua D. Lothringer

5. Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA

   Jea Adams Redai, Mercedes López-Morales, James Kirk & Diana Powell

6. Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, The Netherlands

   Saugata Barat & Jean-Michel Désert

7. Department of Physics & Astronomy, University of Kansas, Lawrence, KS, USA

Jonathan Brande & Ian J. M. Crossfield

8. Astrophysics Section, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Mario Damiano & Renyu Hu

9. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

Tansu Daylan

10. Space Telescope Science Institute, Baltimore, MD, USA

Néstor Espinoza & Nikolay K. Nikolov

11. Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD, USA

Néstor Espinoza, Lakeisha Ramos-Rosado, David K. Sing, Natalie H. Allen & Zafar Rustamkulov

12. Department of Astronomy, Cornell University, Ithaca, NY, USA

Laura Flagg, Ryan J. MacDonald & Jake D. Turner

13. Carl Sagan Institute, Cornell University, Ithaca, NY, USA

Laura Flagg, Ryan J. MacDonald & Jake D. Turner

14. School of Earth and Planetary Sciences (SEPS), National Institute of Science Education and Research (NISER), Homi Bhabha National Institute (HBNI), Jatani, India

Jayesh M. Goyal

15. Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

Renyu Hu, Julie Inglis, Nicole L. Wallack & Heather A. Knutson

16. Center for Space and Habitability, University of Bern, Bern, Switzerland

Elspeth K. H. Lee

17. Max Planck Institute for Astronomy, Heidelberg, Germany

Thomas Mikal-Evans, Laura Kreidberg & Luigi Mancini

18. Department of Physics, Université de Montréal, Montreal, Quebec, Canada

Pierre-Alexis Roy, Björn Benneke, Jake Taylor & Caroline Piaulet

19. Institute for Research on Exoplanets, Université de Montréal, Montreal, Quebec, Canada

Pierre-Alexis Roy, Björn Benneke, Jake Taylor & Caroline Piaulet

20. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

Natalie M. Batalha, Aarynn L. Carter, Sagnick Mukherjee, Joanthan J. Fortney & Kazumasa Ohno

21. Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA

Jacob L. Bean & Adina D. Feinstein

22. Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, USA

Zachory K. Berta-Thompson

23. European Space Agency, Space Telescope Science Institute, Baltimore, MD, USA

Quentin Changeat

24. Department of Physics and Astronomy, University College London, London, UK

Quentin Changeat

25. NASA Goddard Space Flight Center, Greenbelt, MD, USA

Knicole D. Colón

26. Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA

Daniel Foreman-Mackey

27. School of Physics, Trinity College Dublin, Dublin, Ireland

Neale P. Gibson

28. School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

Michael R. Line & Luis Welbanks

29. University Observatory Munich, Ludwig Maximilian University of Munich, Munich, Germany

Karan Molaverdikhani

30. Exzellenzcluster Origins, Garching, Germany

Karan Molaverdikhani

31. Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA

Sarah E. Moran

32. Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain

Giuseppe Morello & Enric Palle

33. Departamento de Astrofísica, Universidad de La Laguna (ULL), Tenerife, Spain

Giuseppe Morello

34. INAF - Palermo Astronomical Observatory, Palermo, Italy

Giuseppe Morello

35. Space Science Institute, Boulder, CO, USA

Julianne I. Moses

36. Steward Observatory, University of Arizona, Tucson, AZ, USA

Everett Schlawin & Megan Mansfield

37. Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA

David K. Sing

38. Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA

Kevin B. Stevenson & Erin M. May

39. Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

Jake Taylor, Vivien Parmentier, Xianyu Tan & Shang-Min Tsai

40. Indian Institute of Technology Indore, Indore, India

Keshav Aggarwal

41. Centre for Exoplanets and Habitability, University of Warwick, Coventry, UK

Eva-Maria Ahrer & Peter J. Wheatley

42. Department of Physics, University of Warwick, Coventry, UK

Eva-Maria Ahrer, Kevin Heng & Peter J. Wheatley

43. School of Physical Sciences, The Open University, Milton Keynes, UK

Joanna K. Barstow

44. Bay Area Environmental Research Institute, NASA Ames Research Center, Moffett Field, CA, USA

Taylor J. Bell

45. Department of Physics, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Jasmina Blecic

46. Center for Astro, Particle, and Planetary Physics (CAP3), New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Jasmina Blecic

47. School of Physics and Astronomy, University of Leicester, Leicester, UK

Sarah L. Casewell

48. Centre for Exoplanet Science, University of St Andrews, St Andrews, UK

Katy L. Chubb

49. Leiden Observatory, University of Leiden, Leiden, The Netherlands

Nicolas Crouzet & Yamila Miguel

50. INAF - Osservatorio Astrofisico di Torino, Pino Torinese, Italy

Patricio E. Cubillos

51. Space Research Institute, Austrian Academy of Sciences, Graz, Austria

Patricio E. Cubillos

52. Institute of Astronomy, Department of Physics and Astronomy, KU Leuven, Leuven, Belgium

Leen Decin

53. Planetary Sciences Group, Department of Physics, University of Central Florida, Orlando, FL, USA

Joseph Harrington

54. Florida Space Institute, University of Central Florida, Orlando, FL, USA

Joseph Harrington

55. Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Munich, Germany

Kevin Heng

56. Institute for Astrophysics, University of Vienna, Vienna, Austria

Nicolas Iro

57. Department of Astronomy, University of Maryland, College Park, MD, USA

Eliza M.-R. Kempton

58. Department of Physics, Imperial College London, London, UK

James Kirk

59. California Institute of Technology, Pasadena, CA, USA

Jessica Krick

60. Laboratoire d'Astrophysique de Bordeaux, Université de Bordeaux, Pessac, France

Jérémy Leconte

61. Département d'Astronomie, Université de Genève Sauverny, Versoix, Switzerland

Monika Lendl & Dominique J. M. Petit dit de la Roche

62. Department of Astronomy, University of Michigan, Ann Arbor, MI, USA

Ryan J. MacDonald

63. Department of Physics, University of Rome "Tor Vergata", Rome, Italy

Luigi Mancini

64. INAF - Turin Astrophysical Observatory, Pino Torinese, Italy

Luigi Mancini

65. Department of Physics and Astronomy, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK

Nathan J. Mayne

66. SRON Netherlands Institute for Space Research, Leiden, The Netherlands

Yamila Miguel

67. Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Laboratoire Lagrange, Nice, France

Vivien Parmentier

68. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Benjamin V. Rackham

69. Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Benjamin V. Rackham

70. Astronomy Department, Wesleyan University, Middletown, CT, USA

Seth Redfield

71. Van Vleck Observatory, Wesleyan University, Middletown, CT, USA

Seth Redfield

72. Institute of Astronomy, University of Cambridge, Cambridge, UK

Laura K. Rogers

73. Maison de la Simulation, CEA, CNRS, Université Paris-Sud, Université Versailles St Quentin, Université Paris-Saclay, Gif-sur-Yvette, France

P. Tremblin

74. Planetary Science Institute, Tucson, AZ, USA

Miguel de Val-Borro

75. Université de Paris Cité and Université Paris-Est Creteil, CNRS, LISA, Paris, France

Olivia Venot

76. Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA

Xi Zhang

Contributions

All authors played a substantial role in one or more of the following: development of the original proposal, management of the project, definition of the target list and observation plan, analysis of the data, theoretical modelling and preparation of this manuscript. Some specific contributions are listed as follows. N.M.B., J.L.B. and K.B.S. provided overall programme leadership and management. L.A. and H.R.W. led the efforts for this manuscript. D.K.S., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., K.B.S., L.K., M.L.-M., M.R.L., N.M.B., V.P. and Z.K.B.-T. made notable contributions to the design of the programme. K.B.S. generated the observing plan, with input from the team. E.S. and N.E. provided instrument expertise. B.B., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., L.K., M.L.-M., M.R.L., N.M.B. and Z.K.B.-T. led or co-led working groups and/or contributed to important strategic planning efforts, such as the design and implementation of the prelaunch Data Challenges. A.L.C., D.K.S., E.S., N.E., N.P.G. and V.P. generated simulated data for prelaunch testing of methods. L.A., H.R.W., M.K.A., N.E.B. and J.D.L. contributed substantially to the writing of this manuscript, along with contributions in Methods from J.A.R., S.B., M.D., N.E., L.F., J.M.G., D.G., J.I., T.M.-E., P.-A.R. and N.L.W. L.A., H.R.W., M.K.A., J.A.R., S.B., M.D., N.E., L.F., D.G., J.I., T.M.-E., P.-A.R. and N.L.W. contributed to the development of data-analysis pipelines and/or provided the data-analysis products used in this analysis, that is, reduced the data, modelled the light curves and/or produced the planetary spectrum,

with further contributions from J.Brande, T.D. and L.R.-R. J.D.L., N.E.B., J.M.G., E.K.H.L. and R.H. generated theoretical model grids for comparison with data. H.R.W., J.D.L. and N.E.B. generated figures for this manuscript. M.L.-M., K.D.C., N.P.G., L.K., M.L., J.I.M. and E.S. provided substantial feedback to the manuscript, coordinating comments from all other authors. T.D. is a LSSTC Catalyst Fellow, N.H.A. and A.D.F. are NSF Graduate Research Fellows, J.Kirk is an Imperial College Research Fellow, R.J.M., M.M., D.P., J.D.T. and L.W. are NHFP Sagan Fellows and B.V.R. is a 51 Pegasi b Fellow.

Corresponding authors

Correspondence to Lili Alderson or Hannah R. Wakeford.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Norio Narita and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 The throughput and spectral trace for WASP-39 across NRS1 and NRS2.

a, Normalized throughput of NRS1 and NRS2 detectors (as custom produced; see Methods, 'Limb-darkening'), which shows the cutoff at short wavelengths. **b**, 2D spectral images of the trace produced from the ExoTiC-JEDI [V1] reduction before cleaning steps. The aspect ratio has been stretched in the *y* direction to show the structure of the trace over the 32-pixel-wide subarray more clearly. The NRS2 spectral position is slightly offset from that of NRS1, as the NRS2 subarray was moved following commissioning to ensure that the centre of the spectral trace fell fully on the detector and did not fall off the top-right corner[139].

Extended Data Fig. 2 Time-dependent decorrelation parameters.

a, The change in the FWHM of the spectral trace at selected wavelengths. This change does not correspond to any high-gain antenna movements and is attributed to a large mirror-tilt event. These measurements demonstrate that the mirror-tilt event has a wavelength dependence. Changes to the PSF have a larger impact at short wavelengths, as the PSF of the spectrum increases with wavelength[139]. **b**,**c**, The change in the *x*-pixel and *y*-pixel position of the spectral trace as functions of time, respectively. Positional shifts are calculated by cross-correlating the spectral trace with a template to measure sub-pixel movement on the detector. The *y*-position shift clearly shows a link to the mirror-tilt event.

Extended Data Fig. 3 Normalized flux offset of the stellar baseline before and after the tilt event as a function of wavelength for NRS1 and NRS2.

Purple denotes NRS1 and orange denotes NRS2. The normalized flux offset is calculated per pixel by measuring the median flux in the stellar baseline before and after the transit and calculating the difference. These differences are then normalized by the before-transit flux and plotted on a common scale. Overplotted are the data binned to a resolution of 10 pixels to match

our presented transmission spectra (Fig. 2). We also show a linear fit to each detector to better quantify the decreasing tilt flux amplitude with increasing wavelength (NRS1 = $-0.00073374x + 0.00707344$, NRS2 = $-0.00067165x + 0.00588128$).

Extended Data Fig. 4 Normalized root-mean-squared binning statistic for three of the 11 reductions detailed in Methods.

In each subplot, the red line shows the expected relationship for perfect Gaussian white noise. The black lines show the observed noise from each spectroscopic light curve for pipelines 1, 3 and 5. To compare bins and noise levels, values for all bins in each pipeline are normalized by dividing by the value for a bin width of 1.

Extended Data Fig. 5 Comparison between all fitting pipelines performed on the spectroscopic light curves.

a, The underlying grey data points show the standard deviation between all transmission spectra per spectral bin. The black line shows the unweighted mean uncertainty on the transit depth per bin. Spikes in the uncertainties correspond to spectral bins with higher standard deviations, probably because of differences in pixel-flagging or sigma-clipping at the light-curve level. **b**, Gaussian PDFs of the normalized transmission spectrum residuals, showing the mean offset and the spread relative to the weighted-average transmission spectrum. **c**, Histograms of the normalized transmission spectrum residuals aligned to zero by subtracting the mean of the distribution that was used to generate the PDF above. In panels **b** and **c**, the coloured lines and numbers correspond to the fitting pipeline used to obtain each transmission spectrum, as summarized in Extended Data Table 1. The dashed lines correspond to the fitting pipeline results presented in Fig. 2, demonstrating that they are drawn from across the distribution.

Extended Data Fig. 6 Gaussian versus flat-line fits to the residual transmission spectrum for CO_2, H_2O, SO_2 and the 4.56-μm feature.

Shown after all other absorption from the best-fit model is subtracted from the data. Each of the Gaussian fits has a higher Bayesian evidence than the flat-line fits, indicating a detection, although to varying degrees of significance.

Extended Data Fig. 7 Model transmission spectra of WASP-39b with PHOENIX and gCMCRT with varying abundances of SO_2.

Model transmission spectra compared with the observed spectral feature at 4.1 μm in the G395H data. At wavelengths short of 3.95 μm, which is outside the SO_2 band, all models overlap, further suggesting that the data can be explained by the presence of SO_2 in the atmosphere. By interpolating these 10 times solar metallicity models, we find a best-fit SO_2 abundance of 4.6 ± 0.67 ppm. With the best-fit PICASO 3.0 at 3 times solar metallicity, we find an SO_2 abundance of 2.5 ± 0.65 ppm.

Extended Data Table 1 Summary of transit light-curve fitting
Extended Data Table 2 RCTE model grids
Extended Data Table 3 Detection significances

Supplementary information

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the

article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Alderson, L., Wakeford, H.R., Alam, M.K. *et al.* Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H. *Nature* **614**, 664–669 (2023). https://doi.org/10.1038/s41586-022-05591-3

- Received: 13 October 2022
- Accepted: 24 November 2022
- Published: 09 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05591-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Early Release Science of the exoplanet WASP-39b with JWST NIRISS**
  - Adina D. Feinstein
  - Michael Radica
  - Xi Zhang

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRCam**
  - Eva-Maria Ahrer
  - Kevin B. Stevenson
  - Xi Zhang

  Nature (2023)

- **Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM**
  - Z. Rustamkulov
  - D. K. Sing
  - S. Zieba

  Nature (2023)

JWST opens a window on exoplanet skies

- Julia V. Seidel
- Louise D. Nielsen
- Subhajit Sarkar

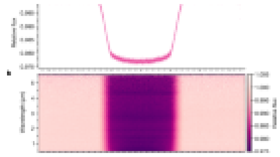
Nature News & Views Forum 15 Feb 2023

A population of red candidate massive galaxies ~600 Myr after the Big Bang

- Ivo Labbé
- Pieter van Dokkum
- Bingjie Wang

Nature Article 22 Feb 2023

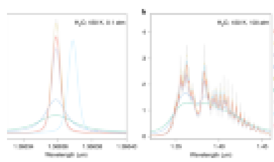
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05591-3



Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM

09 January 2023

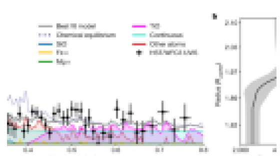
Z. Rustamkulov, D. K. Sing, … S. Zieba



The impending opacity challenge in exoplanet atmospheric characterization

15 September 2022

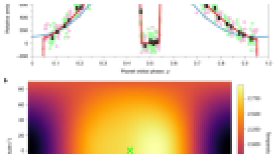
Prajwal Niraula, Julien de Wit, … Roman V. Kochanov



UV absorption by silicate cloud precursors in ultra-hot Jupiter WASP-178b

06 April 2022

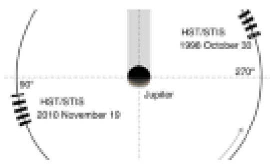
Joshua D. Lothringer, David K. Sing, … Autumn T. Winch



[Diurnal variations in the stratosphere of the ultrahot giant exoplanet WASP-121b]

21 February 2022

Thomas Mikal-Evans, David K. Sing, … Jessica J. Spake



[A sublimated water atmosphere on Ganymede detected from Hubble Space Telescope observations]

26 July 2021

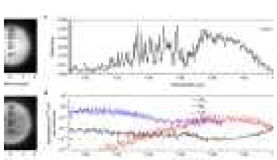
Lorenz Roth, Nickolay Ivchenko, … Kurt D. Retherford



[An absolute sodium abundance for a cloud-free 'hot Saturn' exoplanet]

07 May 2018

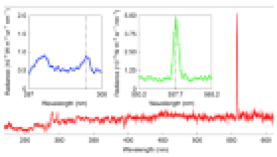
N. Nikolov, D. K. Sing, … J. J. Spake



[Detection of hydrogen sulfide above the clouds in Uranus's atmosphere](#)

23 April 2018

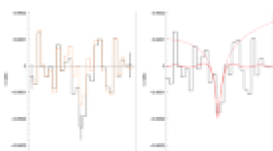
Patrick G. J. Irwin, Daniel Toledo, … Bruno Bézard



[Detection of green line emission in the dayside atmosphere of Mars from NOMAD-TGO observations](#)

15 June 2020

J.-C. Gérard, S. Aoki, … B. Beeckman



[Phosphine gas in the cloud decks of Venus](#)

14 September 2020

Jane S. Greaves, Anita M. S. Richards, … Jim Hoge

- Article
- Open Access
- Published: 09 January 2023

Early Release Science of the exoplanet WASP-39b with JWST

NIRISS

- Adina D. Feinstein ,
- Michael Radica,
- Luis Welbanks,
- Catriona Anne Murray,
- Kazumasa Ohno,
- Louis-Philippe Coulombe,
- Néstor Espinoza,
- Jacob L. Bean,
- Johanna K. Teske,
- Björn Benneke,
- Michael R. Line,
- Zafar Rustamkulov,
- Arianna Saba,
- Angelos Tsiaras,
- Joanna K. Barstow,
- Jonathan J. Fortney,
- Peter Gao,
- Heather A. Knutson,
- Ryan J. MacDonald,
- Thomas Mikal-Evans,
- Benjamin V. Rackham,
- Jake Taylor,
- Vivien Parmentier,
- Natalie M. Batalha,
- Zachory K. Berta-Thompson,
- Aarynn L. Carter,
- Quentin Changeat,
- Leonardo A. dos Santos,
- Neale P. Gibson,
- Jayesh M. Goyal,
- Laura Kreidberg,
- Mercedes López-Morales,
- Joshua D. Lothringer,
- Yamila Miguel,

- Karan Molaverdikhani,
- Sarah E. Moran,
- Giuseppe Morello,
- Sagnick Mukherjee,
- David K. Sing,
- Kevin B. Stevenson,
- Hannah R. Wakeford,
- Eva-Maria Ahrer,
- Munazza K. Alam,
- Lili Alderson,
- Natalie H. Allen,
- Natasha E. Batalha,
- Taylor J. Bell,
- Jasmina Blecic,
- Jonathan Brande,
- Claudio Caceres,
- S. L. Casewell,
- Katy L. Chubb,
- Ian J. M. Crossfield,
- Nicolas Crouzet,
- Patricio E. Cubillos,
- Leen Decin,
- Jean-Michel Désert,
- Joseph Harrington,
- Kevin Heng,
- Thomas Henning,
- Nicolas Iro,
- Eliza M.-R. Kempton,
- Sarah Kendrew,
- James Kirk,
- Jessica Krick,
- Pierre-Olivier Lagage,
- Monika Lendl,
- Luigi Mancini,
- Megan Mansfield,
- E. M. May,
- N. J. Mayne,

- Nikolay K. Nikolov,
- Enric Palle,
- Dominique J. M. Petit dit de la Roche,
- Caroline Piaulet,
- Diana Powell,
- Seth Redfield,
- Laura K. Rogers,
- Michael T. Roman,
- Pierre-Alexis Roy,
- Matthew C. Nixon,
- Everett Schlawin,
- Xianyu Tan,
- P. Tremblin,
- Jake D. Turner,
- Olivia Venot,
- William C. Waalkes,
- Peter J. Wheatley &
- …
- Xi Zhang

Show authors

Nature **volume 614**, pages 670–675 (2023)

- 2486 Accesses

- 1 Citations

- 19 Altmetric

- Metrics details

Abstract

The Saturn-mass exoplanet WASP-39b has been the subject of extensive efforts to determine its atmospheric properties using transmission spectroscopy[1,2,3,4]. However, these efforts have been hampered by modelling degeneracies between composition and cloud properties that are caused by

limited data quality[5,6,7,8,9]. Here we present the transmission spectrum of WASP-39b obtained using the Single-Object Slitless Spectroscopy (SOSS) mode of the Near Infrared Imager and Slitless Spectrograph (NIRISS) instrument on the JWST. This spectrum spans 0.6–2.8 μm in wavelength and shows several water-absorption bands, the potassium resonance doublet and signatures of clouds. The precision and broad wavelength coverage of NIRISS/SOSS allows us to break model degeneracies between cloud properties and the atmospheric composition of WASP-39b, favouring a heavy-element enhancement ('metallicity') of about 10–30 times the solar value, a sub-solar carbon-to-oxygen (C/O) ratio and a solar-to-super-solar potassium-to-oxygen (K/O) ratio. The observations are also best explained by wavelength-dependent, non-grey clouds with inhomogeneous coverageof the planet's terminator.

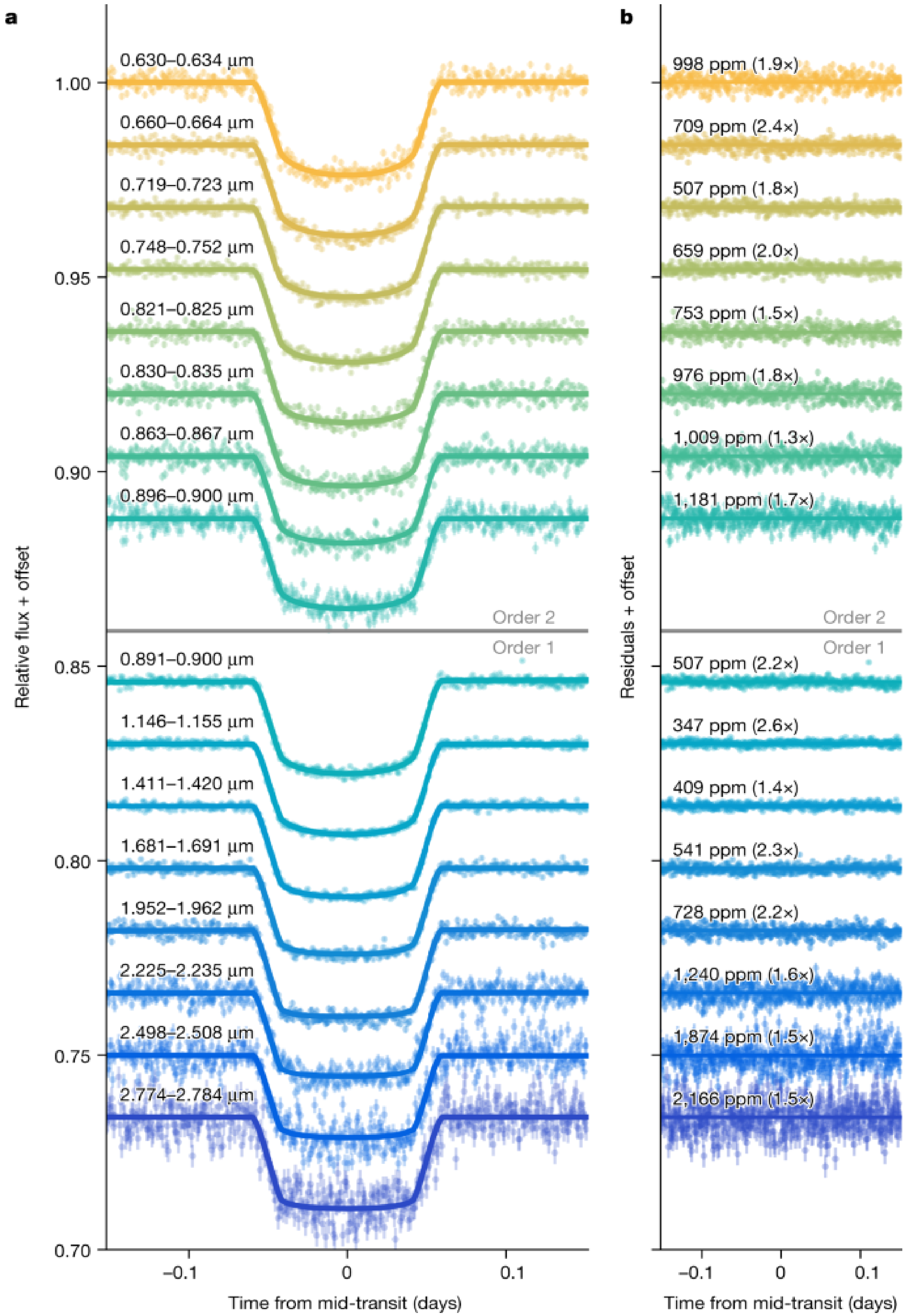
Main

We observed a transit of WASP-39 b using the NIRISS[10] on the JWST as part of the Transiting Exoplanet Community Early Release Science Program[11,12]. Our observations spanned 8.2 h starting on 26 July 2022 20:45 UTC, covering the 2.8-h transit as well as 3.0 h before and 2.4 h after the transit to establish a flux baseline. The data were taken in the SOSS mode, which simultaneously covers the wavelength range from 0.6 to 2.8 μm across two spectral orders on the same detector. Order 1 contains the spectral range between 0.6 and 2.8 μm at an average resolving power of $R \equiv \lambda/\Delta\lambda$ = 700, whereas order 2 delivers the spectral range of 0.6–1.4 μm at an average resolving power of R = 1,400. In the SOSS mode, the spectra are spread across more than 20 pixels in the cross-dispersion direction by means of a cylindrical defocusing lens (see Extended Data Fig. 1), thus allowing longer integration times and reducing the impact of pixel-level differences in the detector response. However, this defocus results in the physical overlap of both orders on the detector. The time-series observation was composed of 537 integrations of 49.4 s (nine groups per integration), corresponding to a duty cycle of 89%.

We extracted the stellar spectra from the time-series observations using six different pipelines to test the impact of differences in spectral order tracing, $1/f$ noise correction, background removal and spectrum extraction

methodology (see Methods and Extended Data Figs. 2 and 3). We created spectrophotometric light curves for each pipeline (Fig. 1) and summed the data to create white-light curves per spectral order (Extended Data Fig. 4). The spectrophotometric and white-light curves are largely free of instrumental systematics except for a constant-rate linear trend in time and an exponential ramp effect within the first 15 min of the time series. The fitted transit depths were binned into 80 spectral wavelength changes in order 1 and 20 in order 2 to create transmission spectra at $R \approx 300$. We present the spectra from the nirHiss, supreme-SPOON and transitspectroscopy reduction pipelines in Fig. 2. We find consistent results between the pipelines, with the derived spectra also being in agreement with previous Hubble Space Telescope (HST) observations (see also Extended Data Fig. 5).

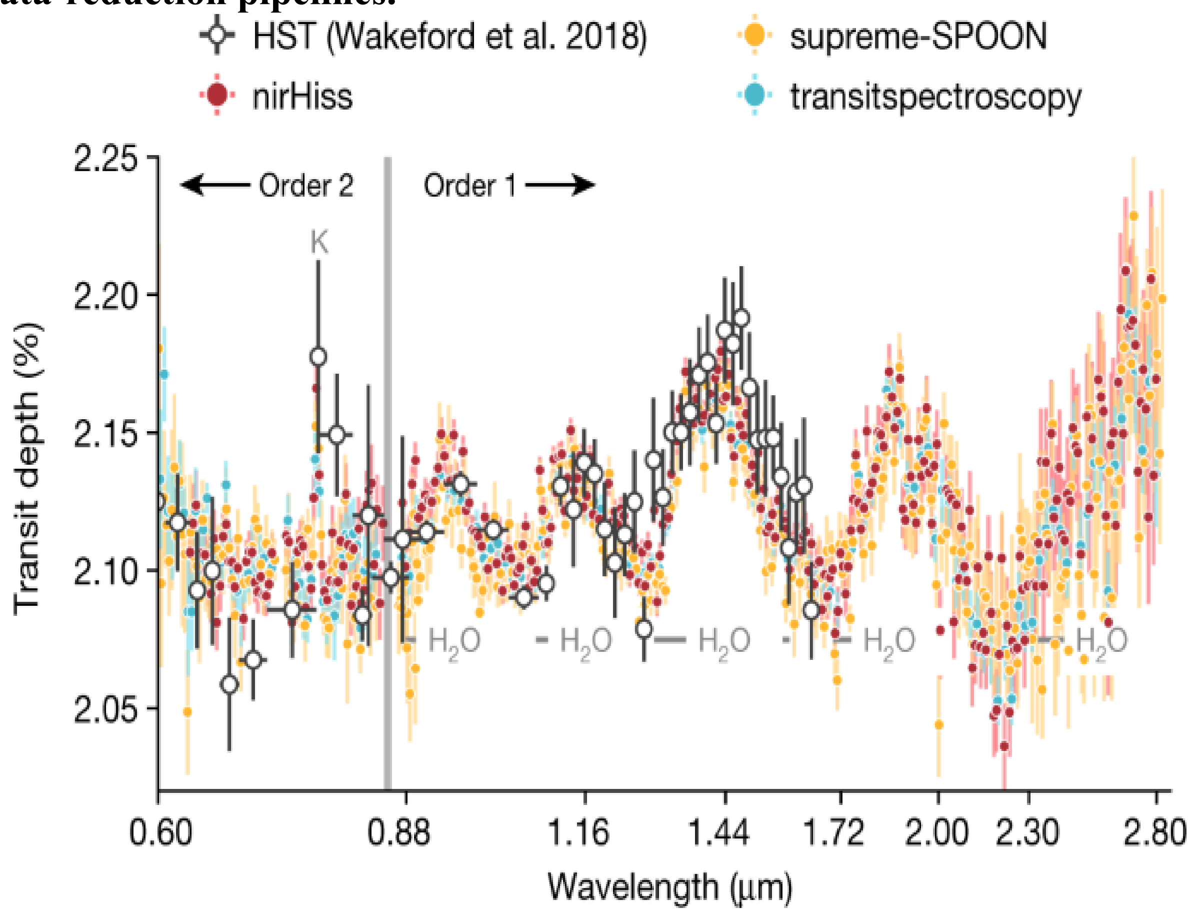
Fig. 1: Selection of systematics-corrected spectrophotometric light curves and residuals for the transit of WASP-39b observed with NIRISS/SOSS for orders 1 and 2.



a
b
0.630–0.634 µm
0.660–0.664 µm
0.719–0.723 µm
0.748–0.752 µm
0.821–0.825 µm
0.830–0.835 µm
0.863–0.867 µm
0.896–0.900 µm
Order 2
Order 1
0.891–0.900 µm
1.146–1.155 µm
1.411–1.420 µm
1.681–1.691 µm
1.952–1.962 µm
2.225–2.235 µm
2.498–2.508 µm
2.774–2.784 µm
1.00
0.95
0.90
0.85
0.80
0.75
0.70
Relative flux + offset
–0.1
0
0.1
Time from mid-transit (days)
998 ppm (1.9×)
709 ppm (2.4×)
507 ppm (1.8×)
659 ppm (2.0×)
753 ppm (1.5×)
976 ppm (1.8×)
1,009 ppm (1.3×)
1,181 ppm (1.7×)
Order 2
Order 1
507 ppm (2.2×)
347 ppm (2.6×)
409 ppm (1.4×)
541 ppm (2.3×)
728 ppm (2.2×)
1,240 ppm (1.6×)
1,874 ppm (1.5×)
2,166 ppm (1.5×)
Residuals + offset
–0.1
0
0.1
Time from mid-transit (days)

An exoplanet transit model (solid line) was fitted to each light curve with chromatic_fitting using a quadratic limb-darkening law. The limb-darkening coefficients, planet-to-star radius ratio (R_p/R_*) and out-of-transit flux were varied in each wavelength channel, whereas all other parameters were fixed. The residuals to the best-fit models are shown for each light curve. The wavelength range for each channel is denoted in panel **a**, whereas parts-per-million (ppm) scatter in the residuals is denoted in panel **b**. We calculate the ppm as the standard deviation of the out-of-transit residuals. We quote the ratio of the predicted photon noise for each bin in brackets. The reductions are from the nirHiss and chromatic_fitting routines described in Methods. We define our errors as the 1σ uncertainties extracted from the stellar spectra.
(https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/figure1.py).

Fig. 2: NIRISS transmission spectra for WASP-39b obtained by three data-reduction pipelines.

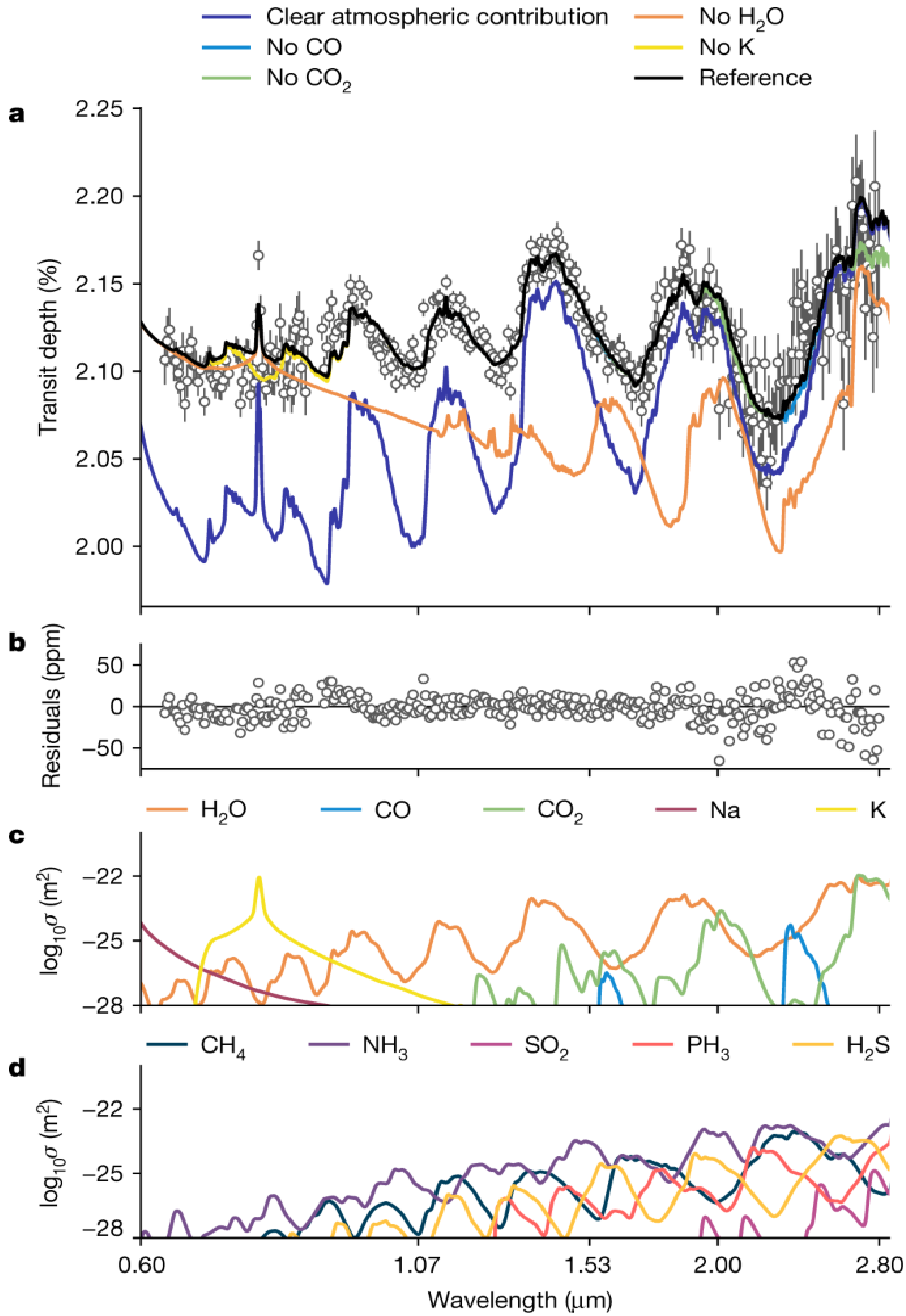


We find broad agreement in the overall structure of the transmission spectra between several reduction pipelines, a sample of which are presented here (see Extended Data Fig. 5 for all reductions). The JWST data are shown in the coloured points, whereas previous HST observations of WASP-39b (ref. [18]) are shown in white. We note that we only consider wavelengths <0.85 µm for order 2, as order 1 has much higher fidelity in the overlapping 0.85–1.0-µm range. We define our errors as the 1σ uncertainties extracted from the 16th and 84th percentiles of the transit depths fit from each pipeline. The JWST and HST data agree across the three broad H_2O features that they have in common. We find evidence of a K absorption feature at 0.76 µm in the new JWST data, which was proposed in the previous HST data[18]. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/figure2.py).

We investigated the atmospheric properties of WASP-39b by comparing our measured transmission spectrum from the nirHiss pipeline to grids of one-dimensional, radiative–convective–thermochemical equilibrium models. These models explore the impact of atmospheric metallicity (M/H), carbon-to-oxygen ratio (C/O), potassium-to-oxygen ratio (K/O), heat redistribution (*f*) and cloud coverage on the transmission spectrum of the planet. We explored several cloud models ranging from parametric treatments[13,14] to a droplet sedimentation model[15] that calculates the vertical distributions of cloud mass mixing ratio and mean particle size from the balance between gravitational sedimentation and eddy diffusion of cloud particles. Using a Bayesian inference framework (see Methods), we compared these grids of models to the observations and inferred the ranges of M/H, C/O ratio, K/O ratio and *f* that best explain the data while marginalizing over different cloud treatments. WASP-39, the host star, has a metallicity equal to that of the Sun within measurement precision[16,17,18,19], so we reference the atmospheric abundances of the planet to the solar pattern of elemental abundances[20]. We compared the grid spectra computed by various models (PICASO, ATMO, PHOENIX and ScCHIMERA) with an observational spectrum obtained from each data-reduction pipeline and obtained broadly consistent results on the inferred atmospheric properties. We report the results from the comparison between the nirHiss spectrum and ScCHIMERA grid that allows the most comprehensive treatments of cloud properties.

Our best-fitting model to the NIRISS/SOSS transmission spectrum of WASP-39 b is presented in Fig. 3. The spectral maxima at 0.9, 1.15, 1.4 and 1.8 µm owing to water absorption result in a >30σ detection of the molecule (see Methods). Similarly, the potassium doublet at 0.768 µm is detected in the data at 6.8σ. Signatures of CO and/or CO_2 are identified because of their contribution to the spectrum past 2.3 µm. We find a 3.6σ significance model preference for CO and no significant preference for CO_2 (see Methods).

Fig. 3: Interpretation of the constituents of the NIRISS WASP-39b transmission spectrum.

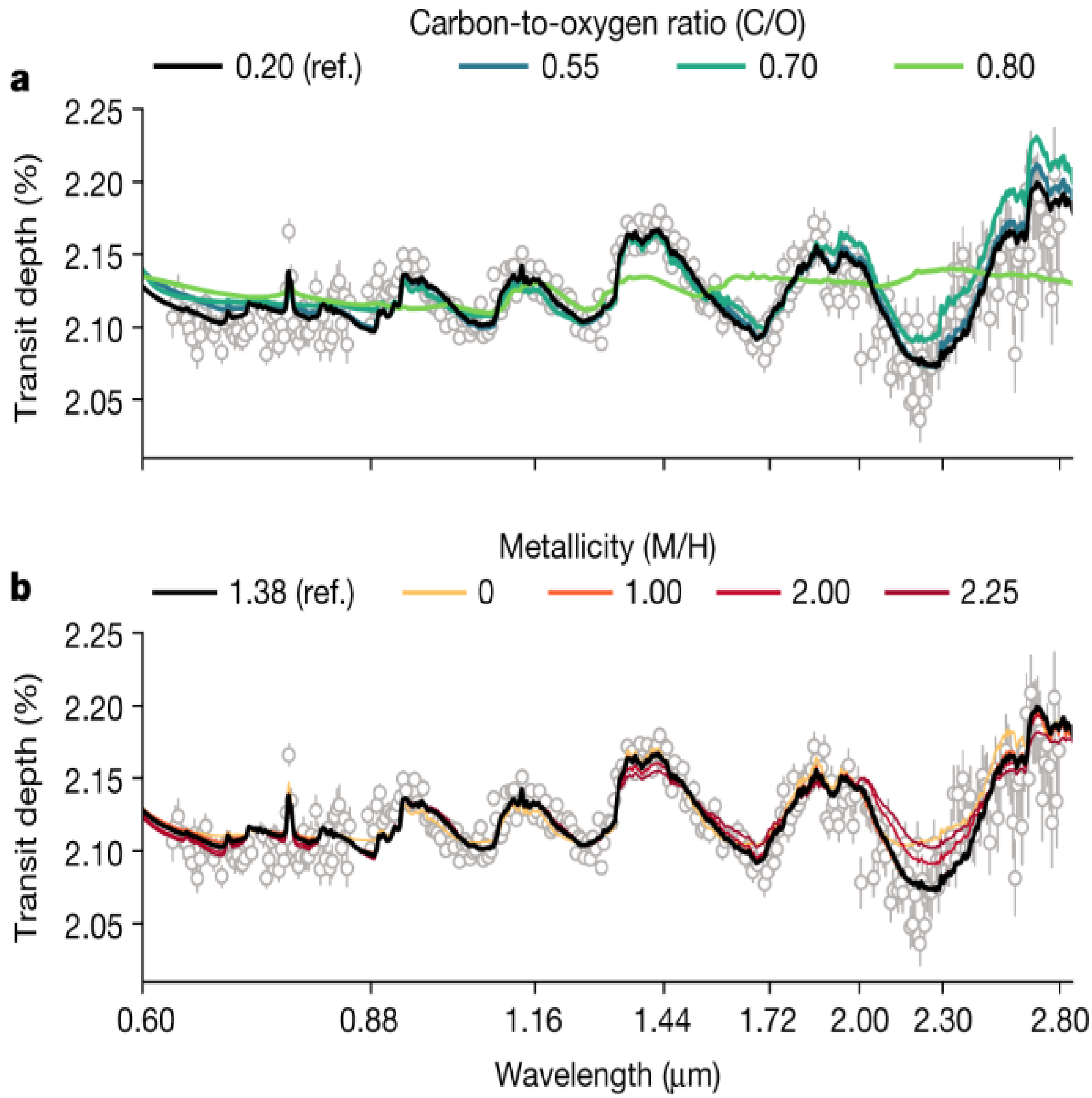


Clear atmospheric contribution
No CO
No CO2
No H2O
No K
Reference
a
Transit depth (%)
2.25
2.20
2.15
2.10
2.05
2.00
b
Residuals (ppm)
50
0
−50
H2O
CO
CO2
Na
K
c
log10σ (m2)
−22
−25
−28
CH4
NH3
SO2
PH3
H2S
d
log10σ (m2)
−22
−25
−28
0.60
1.07
1.53
2.00
2.80
Wavelength (μm)

a,**b**, Panel **a** shows the comparison of the transmission spectrum of WASP-39b from the nirHiss reduction (grey points) with respect to the best-fit reference model (black line). This model assumes an atmospheric metallicity of M/H = 1.38 (23 times the solar value), C/O ratio of 0.2 (0.55 times the solar value[20]), K/O ratio of 0.1 (1.26 times the solar value), full day–night heat redistribution ($f = 1$) and flux-balanced clouds with inhomogeneous terminator coverage. Each coloured line removes a key constituent found in our best-fit reference model to demonstrate how the spectrum would change were these features not included. The removal of clouds and H_2O absorption from the reference model result in large-scale changes to the shape and depth of the transmission spectrum. Other sources of opacity with an impact on the spectrum are K, CO and CO_2. Residuals between the data and the reference model are plotted in panel **b**. **c**,**d**, These two panels show the molecular absorption cross-sections for a selection of gases observable within the NIRISS bandpass. Panel **c** highlights gases inferred by our analysis of the spectrum of WASP-39b. Panel **d** highlights some gases that were not identified in these data but may be present in future observations of other exoplanets. We define our errors as the 1σ uncertainties extracted from the 16th and 84th percentiles of the transit depths fit from each pipeline. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/figure3.py).

From the chemical equilibrium models considered, we find that the observations are best explained by a sub-solar C/O ratio (see Fig. 4a). Across the different spectroscopic resolutions and atmospheric models, the best-fit C/O ratio is 0.2, which is the lowest ratio explored in the grid of models. We rule out super-solar C/O ratio because of the lack of CH_4 features at about 1.7 µm and about 2.3 µm, at which they would be expected for C/O ratio ≳ 0.7. Overall, solar-to-super-solar C/O ratios fail to explain the transmission spectrum at the shortest (≲1 µm) and longest (≳2 µm) wavelengths. Our best-fit C/O ratio is broadly consistent with the observations of WASP-39b with NIRCam (2.4–4.0 µm; ref. [21]), NIRSpec G395H (3–5 µm; ref. [22]) and NIRSpec PRISM (0.5–5.0 µm; ref. [23]).

Fig. 4: Impact of the C/O ratio and metallicity on the JWST-NIRISS spectrum of WASP-39b.



a, Variation of the C/O ratio in the best-fit reference model while keeping the metallicity, redistribution and K/O ratio parameters from the reference model the same and fitting for the cloud parameters and scaled planetary radius to best explain the observations. Under these equilibrium conditions, increasing the C/O ratio results in less H_2O and more CH_4, the latter having spectroscopic signatures incompatible with the observations. To mute these incompatible CH_4 features at high C/O ratios, the model requires a higher degree of cloudiness that also mutes any remaining H_2O features in the spectrum. **b**, The same as for panel **a** but instead we vary the metallicity parameter. The metallicity constraint is driven by the $\lambda > 2$ μm data; the

high-metallicity models (M/H > 2) expect larger transit depths than that seen in the data. The same reference model is plotted as a thick black line in both panels. We define our errors as the 1σ uncertainties extracted from the 16th and 84th percentiles of the transit depths fit from each pipeline. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/figure4.py).

We find that the observations are best explained by an atmospheric metallicity of 10–30 times solar. Metallicity inferences over the wavelength range of these observations are largely driven by the size and shape of the water-vapour features, with some minor contributions because of CO and/or CO_2 at longer wavelengths (>2 µm; see Figs. 3 and 4b). The preferred range of metallicities provides the best fit to the shape and size of the muted water-vapour features shortward of 2 µm in combination with the larger water and CO/CO_2 feature longward of 2 µm, regardless of the assumed cloud treatment in our models.

Owing to the simultaneous detection of potassium and water vapour, we are able to place constraints on the K/O ratio, which is a refractory-to-volatile elemental ratio, being a solar-to-super-solar value. Because the refractory elements are condensed into solids in most parts of protoplanetary disks, the disk gas accretion tends to cause a sub-stellar refractory elemental abundance[24]. By contrast, solid accretion, such as planetesimal accretion, acts to increase the refractory elemental abundance and refractory-to-volatile elemental ratio[25], although the latter depends on the composition of the accreted solids[26]. We anticipate that the K/O ratio diagnoses to what degree the solid accretion enriched the atmosphere during the formation stage. All of our fitted models find that the WASP-39b observations are well described by solar-to-super-solar K/O ratios, which is in agreement with previous inferences for this planet obtained through observations with limited spectral coverage[27]. We do not expect the K feature to be affected by stellar chromospheric magnetic activity given the effective temperature of the star of approximately 5,300 K (ref. [28]) and the general quietness of WASP-39 (see ref. [21]). It is also in line with larger population studies of hot giant planets that broadly found solar-to-super-solar refractory abundances and solar-to-sub-solar H_2O abundances[27,29]. The shape and strength of the potassium doublet are best explained by K/O ratios of 0.1–0.5, equivalent to

1–3 times solar (see Extended Data Fig. 8), whereas the suggested K/O ratio might be a lower limit owing to the photoionization of K at upper atmospheres[30].

The NIRISS/SOSS observations enable the detection of clouds in the atmosphere of WASP-39b. Clear atmosphere models cannot explain the amplitudes of all of the water-vapour features simultaneously, which strongly indicates the presence of clouds (see Methods and Extended Data Fig. 6). The atmospheric models explored here indicate the presence of non-grey and non-homogeneous clouds, with model preferences of 8σ and greater for models with both non-grey and non-homogeneous clouds over models with grey homogeneous clouds only. This model preference is driven by the decrease in transit depth between 2.0 and 2.3 μm (see Extended Data Fig. 7a), which cannot be explained by grey clouds uniformly distributed along the terminator (see Extended Data Fig. 7a). Moreover, in the various cloud treatments tested here (grey, grey + power law and flux-balanced clouds; see Methods), both parametric and droplet sedimentation models indicate a preference for inhomogeneous cloud coverage of roughly 50–70% around the planetary day–night terminator because it better explains the decrease in transit depth between 2.0 and 2.3 μm.

Atmospheric circulation and cloud microphysical models have predicted that the cloud structure varies substantially along the terminators of hot Jupiters[31,32,33]. In particular, different compositions of clouds have different condensation temperatures and thus probably have different cloud coverage at the terminator[31]. Further studies combining temperature difference of east–west terminators to microphysical cloud models may be able to use the measured cloud coverage to determine the cloud composition of WASP-39b. Previous indications of non-grey or non-homogeneous clouds[34,35,36,37,38,39] have relied on a single or small number of spectroscopic points, making our inference here for WASP-39b of non-grey cloud with inhomogeneous terminator coverage in the transmission spectrum of an exoplanetary atmosphere the most confident so far. These constraints on the physical properties of clouds, alongside several spectral features across a broad wavelength coverage, are key to breaking well-known degeneracies between the metallicity and cloud cover in atmospheric models[8,14,40] and deriving constraints on the bulk atmospheric properties.

The high precision of NIRISS/SOSS in combination with broadest wavelength coverage <2.8 μm for any JWST instrument, minimal systematics and no issue with saturation allows us to obtain more precise and robust constraints on atmospheric composition and tracers of planet formation than most previous transmission spectroscopy observations. The super-solar metallicity of WASP-39 b and the solar-to-super-solar K/O ratio are in agreement with previous studies of mass-metallicity trends in transiting exoplanets[3,7,27,41]. If confirmed with further detailed modelling, a super-solar K/O ratio in the atmosphere of WASP-39 b would probably indicate enrichment resulting from the accretion of planetesimals[25,26,27], although the measurements of potassium and oxygen abundances for the host star are also needed to establish this result. Similarly, the suggestion of sub-solar C/O ratio and super-solar metallicity may be compatible with a planetesimal accretion scenario, for example, refs. [42,43,44]. The combination of a super-solar metallicity, super-solar K/O ratio and sub-solar C/O ratio may suggest that the planet formed beyond the H_2O snow line followed by inward migration, for which theory predicts efficient accretion of planetesimals at approximately 2–10 AU, for example, refs. [25,45]. At those orbital distances, the planetesimals probably contain K rock (for example, alkali feldspar $KAlSi_3O_8$ (refs. [46,47])) and H_2O ice but almost no CO ice, for example, refs. [48,49], which explains the sub-solar C/O ratio and super-solar K/O ratio, along with a super-solar metallicity if a sufficient amount of planetesimals was accreted. However, fully understanding the possible formation pathways of this planet requires statistical constraints on the complete chemical inventory of the planet and the relative abundances of the carbon-bearing, oxygen-bearing and alkali-bearing species. Such efforts will be possible when applying retrieval techniques to the complete transmission spectrum of WASP-39b from 0.5 to 5.5 μm that is being produced by the Transiting Exoplanet Community Early Release Science Program. Our results validate the JWST's NIRISS/SOSS as an instrument mode fully capable of producing excellent exoplanet atmosphere measurements.

Methods

Given the newness of the data, we applied six independent data-reduction and light-curve-fitting routines to the data: nirHiss, supreme-SPOON,

transitspectroscopy, NAMELESS, iraclis and FIREFLy. Each pipeline extracts the stellar spectra from orders 1 and 2 ($\lambda = 0.6$–2.8 μm) with the exception of FIREFLy, which only extracts data from order 1. There is an extra order 3 that has a spectral range of $\lambda = 0.6$–0.95 μm (ref. [50]). However, the signal of order 3 is generally weak and, because it provides no new wavelength information beyond what is covered in orders 1 and 2, is not used by any of the presented pipelines. Below, we first describe the important reduction steps taken by each, followed by their light-curve-fitting methodologies. We note here that, in each pipeline, the position of the SOSS trace was found to match almost perfectly with that measured during commissioning (Fig. 1). Furthermore, each pipeline trimmed the first 10–15 integrations to remove the effects of the exponential ramp in the fitting routines. We present a summary of all pipelines in Extended Data Table 1.

The nirHiss pipeline

nirHiss is a Python open-source package that uses the stage 2 outputs from the Eureka! pipeline and performs further background and cosmic-ray removal, as well as extraction of the stellar spectra. For this analysis, we took the uncalibrated images and ran our own stages 1 and 2 calibration using Eureka![51], an open-source package that performs spectral extraction and light-curve fitting for several JWST instruments. We use the default steps presented in Eureka!, which includes detector-level corrections, production of count-rate images, application of physical corrections and calibrations to individual exposures.

Next, nirHiss removes background noise sources in a multistep process. The zodiacal background is first removed by applying the background model provided on the STScI JDox User Documentation website (https://jwst-docs.stsci.edu/). The background is scaled to a small region of each science integration in which there was no contamination from any of the orders; in this case, $x \in [190, 250]$, $y \in [200, 500]$. The average scaling—calculated here to be 0.881—is applied to all science integrations. Second, a model of zeroth-order contaminants is built using the F277W integrations. The F277W integrations were taken after the transit of WASP-39b with the GR700XD/CLEAR pupil element and the F277W filter (throughput centred at $\lambda = 2.776$ μm with a bandwidth of $\lambda = 0.715$ μm). These observations

consist of ten integrations with an exposure time of 49.4 s. Observations with the F277W filter contain only the spectral trace of order 1 in the region in which $x \leq 460$ pixels, thus allowing for the detection and modelling of zeroth-order contaminants across most of the detector. A median F277W frame is created to identify and mask any bad-data-quality pixels.

To ensure that no further noise is added from the median F277W frame, we create a 2D background model map using photutils.Background2D. To identify regions of the background, we masked the upper-left corner, in which the trace is located, and any regions >1.5σ, which includes the zeroth-order sources. For photutils.Background2D, we used a filter size of (3, 2) pixels and a box size of (2, 2) pixels. Once the background is removed from the median F277W frame, we apply a Gaussian filter with a width of 2 to smooth out any further small-scale background noise. To apply the median F277W frame to the stage 2 science integrations, we scaled it to two isolated zeroth-order sources in the science integrations at $x_1 \in [900, 1{,}100]$, $y_1 \in [150, 250]$ and $x_2 \in [1{,}800, 2{,}000]$, $y_2 \in [150, 250]$. We applied the average scaling to all integrations. We found the average F277W background scaling to be 2.81. We apply the scaled background frame to each time-series observation integration.

Once the zeroth-order contaminants are removed, we trace the location of orders 1 and 2. The spatial profile for NIRISS/SOSS along the column is double-peaked, with a slight dip in the middle. We developed a routine to identify the trace locations using a three-step approach to identifying each order. For each column in the first order trace, we identify the locations of the two peaks, or ‘ears’, and assume that the middle of the trace is the median row pixel between the two ears. We repeat this process for the third and second orders in that sequence, masking orders once they have been traced. We chose to identify the third order before the second order because it is better spatially resolved and does not overlap with any other orders. The routine creates one main set of traces from a median frame of all observations, which is used to extract the stellar spectra. As an extra output, we track the changes in the (x, y) pixel positions of each order on the detector across all integrations.

After the traces are identified, we continue our reduction to remove any extra noise and cosmic rays/bad pixels. We perform further 1/*f* noise correction following the routine presented in transitspectroscopy (described below). Finally, nirHiss identifies and interpolates over cosmic rays. To identify cosmic rays, we used the L.A.Cosmic technique wrapped into ccdproc[52,53], which identifies pixels based on a variation of the Laplacian edge detection. We identify cosmic rays as pixels with $\sigma > 4$ using this method. We interpolate over any further bad pixels by taking the median value of the two surrounding pixels along the column. We extract the spectra using a box-extraction routine and ignore any contaminants from overlapping orders or from any potential background orders. We use a box diameter of 24 pixels for both orders 1 and 2.

The supreme-SPOON pipeline

In parallel, we reduce the WASP-39 b time-series observation with the independent supreme-SPOON (supreme-Steps to Process sOss ObservatioNs) pipeline, which processes SOSS time-series observations from the raw, uncalibrated detector images to extracted 1D light curves. An outline of the specific steps is presented below.

For detector-level processing, supreme-SPOON closely follows stage 1 of the jwst pipeline. All default steps, up to and including the reference pixel correction, are run using their default settings. The reference pixel step is known to provide an inadequate correction of 1/*f* noise for SOSS observations; however, we include it to remove group-to-group variations in the bias level, as well as even–odd row variations. At this stage, we remove the zodiacal background from each group. This is accomplished by first calculating a group-wise median frame and scaling the model background provided in the STScI JDox to the flux level of each group in this median, yielding eight background models, one for each group. The region chosen to calculate the scaling was $x \in [300, 500]$, $y \in [210, 250]$, in which there is minimal contamination from any of the SOSS orders. The *n*th background model is then subtracted from the corresponding group of each integration.

We then proceed to a more in-depth treatment of 1/*f* noise. Unlike the other pipelines used in this work, supreme-SPOON treats 1/*f* noise at the group

level instead of at the integration level. 1/*f* noise is a time-varying noise source introduced by the voltage amplifiers during the readout of the detector and therefore the 1/*f* pattern will vary from group to group, even within a given integration. To perform the 1/*f* correction, first a median out-of-transit frame is calculated for each group. This group-wise median is then scaled to the flux level of each frame in a given group by means of the transit curve and subtracted, showing the characteristic 1/*f* striping in the residuals. A column-wise median of this residual map is then subtracted from the original frame. The trace residuals as well as any bad pixels are masked in the median calculation.

From this point, we once again proceed with the standard stage 1 steps of the jwst pipeline, with the exception of the dark current step, to obtain the supreme-SPOON stage 1 outputs. The dark current subtraction step is skipped as it was found to reintroduce 1/*f* noise into the data. The dark current level is also extremely small (several tens of electrons s^{-1} compared with many thousands for the target signal) and can thus be safely ignored. supreme-SPOON only applies the assign_wcs, srctype and flat_field steps of the stage 2 jwst pipeline to the stage 1 products. The background subtraction was already performed as part of stage 1 calibrations. Furthermore, the flux calibration steps (pathloss, which accounts for light incident on the telescope primary mirror that falls outside the SUBSTRIP256 subarray, and photom, which performs the actual photometric flux calibration) are skipped, both because an absolute flux calibration is unnecessary for relative spectrophotometric measurements and a wavelength-dependent flux calibration is nonsensical for SOSS, in which contributions from several wavelengths from all orders affect a single pixel. At this point, supreme-SPOON identifies any remaining hot pixels through median filtering of a median stack of all frames and interpolates them by means of the median of a surrounding box. These products are the supreme-SPOON stage 2 results.

Stage 3 of the supreme-SPOON pipeline is the 1D extraction. This can be performed through two different methods: the first is a simple box aperture extraction on each order, ignoring the order contamination. The second uses ATOCA (Algorithm to Treat Order ContAmination)[50] to explicitly model the order contamination. Briefly, ATOCA constructs a linear model for each pixel on the detector, including contributions from the first and second diffraction orders, allowing for the decontamination of the SOSS detector—

that is, ATOCA constructs models of both the first and second orders individually, thereby allowing a box extraction to be performed on each free from the effects of order contamination. Although the effects of this order contamination for differential measurements (such as exoplanet atmosphere observations) are predicted to be small (about 1% of the amplitude of the expected spectral features)[50,54], in the quest to obtain the most accurate possible transmission spectra, this contamination effect is important to take into account. ATOCA is at present built into the Extract1dStep of the jwst pipeline, although it is not the default option and must be toggled to on by means of the 'soss_atoca' parameter. To improve the performance of ATOCA, we do not use the default specprofile reference file included in the jwst pipeline but instead construct estimates of the underlying spatial profiles of the first and second orders, on which ATOCA relies, using the APPLESOSS (A Producer of ProfiLEs for SOSS) algorithm[54]. We determine the centroid positions for each order on a median stack using the 'edgetrigger' algorithm[54], and these positions are found to match to within a pixel with the default centroids contained in the jwst_niriss_spectrace_0023.fits reference file; the spectrace file is available on the JWST Calibration Reference Data System (CRDS). The SOSS trace position is furthermore highly stable over the course of this time-series observation, with root mean square (RMS) variations in x and y positions of approximately 5 mpix and RMS rotation of about 0.3″. We therefore fix the 'soss_transform' parameter to [0, 0, 0] and perform the extraction with a box size of 25 pixels. Any remaining $>5\sigma$ outliers in the resulting spectra are then identified and clipped. At present, supreme-SPOON does not explicitly treat contamination from zeroth orders of background stars that intersect the trace.

The transitspectroscopy pipeline

This third pipeline analysis combines the jwst pipeline stage 1 'rateints.fits' files with transitspectroscopy[55]. transitspectroscopy completes stellar spectral extraction as well as transit fitting.

The trace positions for NIRISS orders 1 and 2 were determined using transitspectroscopy.trace_spectrum. This routine cross-correlates an input function with each column in the detector to find the centre of the different

traces by means of the maximum of the resulting cross-correlation function. To follow the shape of the NIRISS order profiles, an input function consisting of a double Gaussian was used with parameters that were trained on the NIRISS/SOSS observations of HAT-P-14 b (JWST Program ID 1541; principal investigator Espinoza): $\mu_1 = -7.5$; $\sigma_1 = 3.0$; $\mu_2 = 7.5$; $\sigma_2 = 3.0$. The trace for order 2 was not fit for pixels ≤1,040, as the throughput is not high enough for the method to robustly fit the trace without incorporating nearby contaminants. After identifying the trace positions with this method for both orders 1 and 2, both traces were smoothed using a series of spline functions. We find that the best-fit parameters for order 1, which were trained on the HAT-P-14 b observations, are: $x_{\mathrm{knots},1}$ = [[6, 1,200-5], [1,200, 1,500-5], [1,500, 1,700-5], [1,700, 2,041]]; $n_{\mathrm{knots},1}$ = [4, 2, 3, 4] and for order 2: $x_{\mathrm{knots},2}$ = [[601, 850-5], [850, 1,100-5], [1,100, 1,749]]; $n_{\mathrm{knots},2}$ = [2, 2, 5].

The zodiacal background was removed by scaling the model background provided on the STScI JDox User Documentation. This model was compared with a small region of the median science integrations in which there was little to no contamination from the orders ($x \in [500, 800]$, $y \in [210, 250]$). The ratio of all the pixels in this region versus the pixels in the background model was computed, ordered and the median ratio of all the second quartile pixels was used as the scaling factor between the background model and the data, which was found to be 0.909. All the integrations had this scaled background subtracted.

Each integration is corrected for $1/f$ noise with the following procedure. First, all the out-of-transit, background-corrected integrations are median combined and scaled by the relative flux decrease produced by the transit event at each integration (that is, 1.0 for out-of-transit integrations or about 0.976 for mid-transit). These scaled median frames are then subtracted from each individual integration, which then leaves in the frame only detector-level effects, such as $1/f$ noise. We then go column by column and take the median of all pixels in these residual frames within a distance of 20–35 pixels from the centre of the trace, and use this as an estimate of the contribution from $1/f$ noise to that given column. This value is then removed from each pixel within 20 pixels from the trace on that column. No correction for order 1 contamination on order 2 was made, as the

contribution is negligibly small in this case[50]—similarly for order 1 contamination in order 2 in our extraction.

To extract the resulting background-corrected and $1/f$-corrected spectrum, we used the transitspectroscopy.spectroscopy.getSimpleSpectrum routine with a 30-pixel total aperture for both orders. To handle obvious outliers in the resulting spectrum because of, for example, uncorrected cosmic rays and/or deviating pixels, we median-normalized the spectra for each integration and combined them to form a 'master' 1D spectrum for both orders 1 and 2. The median was taken at each wavelength, as well as the error on that median, and this was then used to search for 5σ outliers on each individual integration at each wavelength. If outliers were found, they were replaced by the rescaled version of this median master spectrum.

The NAMELESS pipeline

Starting from the jwst pipeline stage 1 products, we use the NAMELESS (Niriss dAta reduction MEthod for exopLanEt SpectroScopy) pipeline to go through the jwst pipeline stage 2 with the addition of custom correction routines.

First, we go through the assign_wcs, srctype and flat_field steps of the jwst pipeline stage 2, opting for a custom background-subtraction routine and skipping the pathloss and photom steps as absolute flux calibration is not needed. After flat-field correction, we scale the model background provided on the STScI JDox User Documentation to a region of the median frame in which the contribution from the tail of the three orders is lowest ($x \in [200, 250]$, $y \in [400, 600]$). From the distribution of the scaling values of all pixels within the defined region, we take the 16th percentile as our scaling value and subtract the scaled background frame from all integrations.

We subsequently correct for $1/f$ noise by performing a column-by-column subtraction for each median-frame-subtracted integrations. The median frame is computed from the out-of-eclipse integrations (integration # $\in [200, 400]$) and scaled to each individual integration by dividing the sum of the pixels in the first order by that of the median frame. We then subtract the scaled median frame from all integrations, perform the column-by-column

subtraction on the residual frames and add back the scaled median frame to the corrected residual frames to obtain the $1/f$-corrected integrations.

We detect outliers frame by frame using the product of the second derivatives in the column and row directions. This method works particularly well for isolated outliers, as this leads to a strong inflexion that corresponds to a large second derivative. Because the spectral orders also lead to larger second-derivative values, we divide the frames into windows of 4×4 pixels, compute the local second-derivative median and standard deviation and flag any pixel that is more than four standard deviations away from the median. Furthermore, we also flag pixels with null or negative flux. All identified outliers are set equal to the median value of the window in which it was identified.

Finally, we proceed with spectral extraction of the corrected frames by first tracing the sections of the spectral orders that we wish to extract. We trace orders 1 and 2 from $x_1 \in [4, 2{,}043]$ and $x_2 \in [4, 1{,}830]$, respectively. The centre of the traces is found for each individual column by performing a convolution of the profile with a Gaussian filter, in which we use the maximum of the convolved profile as the centre of the trace. For the tracing of the second order, we keep the centre of the trace fixed below $x = 900$, as the flux from the first order can bias the tracing method. Furthermore, we smooth the positions of the trace centroids using a spline function with 11 and 7 knots for the first and second orders, respectively. We perform spectral extraction of the first and second orders at all integrations using the transitspectroscopy.spectroscopy.getSimpleSpectrum routine with an aperture width of 30 pixels.

The iraclis pipeline

We used the jwst pipeline stage 1 rateints.fits files with modified routines from iraclis[5,56], which was initially designed for the HST. The modified routines will be part of the version 2 of iraclis, which will become publicly available in the near future. The routines applied to the rateints.fits files were flat fielding, bad-pixels and cosmic-rays correction, sky background subtraction, $1/f$ noise correction, X-drift and Y-drift detection, light-curve extraction, light-curve modelling and planetary spectrum decontamination.

We started our analysis by dividing the images by the appropriate flat-field frame (jwst_niriss_flat_0275.fits), as provided by the JWST CRDS. The next step was the bad-pixels and cosmic-rays correction. For bad pixels, we used those with a positive DQ flag in the rateints.fits files, excluding the warm pixels, as their large number did not allow for a reliable correction. We also identified extra outliers (cosmic rays or other artefacts) by calculating two flags for each pixel: the difference from the average of the ten horizontally neighbouring pixels (x-flag) and the difference from the average of the ten vertically neighbouring pixels (y-flag). If a pixel's x-flag was 5σ larger than the other pixels in the column and its y-flag 5σ was larger than the other pixels in the row, it was identified as a cosmic ray (see also ref. [56]). Both bad pixels and outliers were replaced with the value of a 2D interpolation function, created from the rest of the pixels, similarly to analyses with the HST[56].

We then subtracted a column-based sky background frame and a column-based $1/f$ noise frame from each image. For each image, we first used a trace filter (value >0.001 in the jwst_niriss_spectrace_0023.fits, provided by the JWST CRDS) and a column-based $1 \times$ median absolute deviation filter to find the illuminated pixels. Then, we calculated the column-based median of the image—using only the unilluminated pixels—and subtracted it from the image. Finally, we calculated the column-based median of the IMFD (Image-MedianFrame Difference)—using only the unilluminated pixels—and subtracted it from the image. This process is not efficient in subtracting 100% of the background contamination, which was removed during the last analysis step (spectrum decontamination).

X-pixel and Y-pixel trace drifts were detected relative to the first image by comparing the sums along the columns and the rows, respectively, similarly to the HST[56]. The drifts are on the order of pixels without any evident trend in motion. Because this is below the subpixel size used in the iraclis extraction, we find that there is no marked impact of not correcting these drifts. For each spectroscopic image, we initially divided each pixel into a 100×100 grid of subpixels and, for each subpixel, we calculated the distance from the trace (CD) and the wavelength (λ), creating the CD_{map} and the λ_{map}, respectively. λ was assigned to each subpixel directly from the wavelength solution (interpolated wavelength solution from the

jwst_niriss_wavemap_0013.fits file, provided by the JWST CRDS, shifted by the detected X and Y drifts). CD was calculated as the distance between the centre of the subpixel and the point of the trace with the same distance (interpolated trace from the jwst_niriss_spectrace_0023.fits file, provided by the JWST CRDS, shifted by the detected X and Y drifts). Our high-resolution bins had a λ width of 10 Å, ranging between 0.62 and 0.85 µm for order 2 and between 0.85 and 2.8 µm for order 1, and a CD width of 1.5 pixels, ranging from −25 to 25 pixels.

Finally, to construct the light curve of each bin, we applied the following smoothed aperture mask on each spectroscopic image and summed the values of all the subpixels. We chose a smoothed aperture, similarly to the HST to reduce the effects of jitter noise:

$${\rm{MASK}}=0.5\times \left[{\rm{erf}}\left(\frac{\left({{\rm{CD}}}_{{\rm{map}}}-{{\rm{CD}}}_{1}\right)}{{\sigma }_{{\rm{CD}}}\sqrt{2}}\right)-{\rm{erf}}\left(\frac{\left({{\rm{CD}}}_{{\rm{map}}}-{{\rm{CD}}}_{2}\right)}{{\sigma }_{{\rm{CD}}}\sqrt{2}}\right)\right]\times \left[{\rm{erf}}\left(\frac{\left({\lambda }_{{\rm{map}}}-{\lambda }_{1}\right)}{{\sigma }_{\lambda }\sqrt{2}}\right)-{\rm{erf}}\left(\frac{\left({\lambda }_{{\rm{map}}}-{\lambda }_{2}\right)}{{\sigma }_{\lambda }\sqrt{2}}\right)\right]$$

in which CD_1, CD_2 and σ_{CD} are the bin boundaries and the smoothing factor along the cross-dispersion axis, respectively, and λ_1, λ_2 and σ_λ are the bin boundaries and the smoothing factor along the dispersion axis, respectively. For the smoothing factors, we used the values of $\sigma_{CD} = 0.015$ pixels and Å—that is, about 10% of the bin size. We chose these values for the smoothing factors because lower values would effectively create a sharp-edge aperture, whereas larger values would force the bins to overlap substantially.

FIREFLy

Although FIREFLy (Fast InfraRed Exoplanet Fitting for Light curves)[57] was written and optimized for reducing NIRSpec-PRISM and G395H time-series observations, it worked well on the NIRISS/SOSS dataset, in which it selected and processed the spectrophotometry from order 1 only with minimal tuning or intervention. FIREFLy is not written in such a way to extract data from order 2 ($\lambda < 0.9$ μm). In our reduction, we perform standard calibrations on the raw data using the jwst pipeline for stages 1 and 2 reduction. On the jwst stage 2 outputs, we perform bad-pixel and cosmic-ray cleaning on each integration. We perform $1/f$ destriping and background subtraction using a pixel mask generated from the temporally medianed image that selects regions of the data image below a specified count threshold. We extract the spectrophotometry using an optimized aperture extraction width that is constant in wavelength. The aperture width is selected such that the scatter of the resulting out-of-transit white-light photometry is minimized.

Light-curve fitting and transmission spectra

We used a suite of light-curve-fitting routines to fit the extracted light curves. Each routine fits for orbital parameters from the broadband white-light curves for each order (see Extended Data Fig. 8). We fixed the orbital period to the best-fitting value from $P = 4.05528$ days (ref. [1]) for all pipeline fits. For the spectroscopic light curves, most routines (nirHiss/chromatic_fitting, supreme-SPOON/juliet, transitspectroscopy/juliet and NAMELESS/ExoTEP) fixed the orbital parameters (that is, the mid-transit time, t_0, semi-major axis to stellar radius ratio a/R_*, impact parameter b, eccentricity e) to the same values to ensure consistency. These parameters were fixed to their best-fitting values from the transitspectroscopy/juliet white-light-curve fit, except for t_0 which was fixed to the value obtained from the white-light curve in each case. This left the planet-to-star radius ratio R_p/R_*, the limb-darkening coefficients and parameters for any further systematics models to vary. These four routines also fit spectroscopic light curves at the native instrument resolution. However, two routines, iraclis and FIREFLy, instead fixed the orbital parameters in their spectroscopic fits to values obtained through their white-light curve fits. iraclis also fits directly for the orbital inclination, i, as opposed to b and a/R_* like the other routines.

iraclis fits for their spectrophotometric light curves at the pixel resolution, whereas FIREFLy binned the spectroscopic light curves first and fits for the transit parameters. We present all of the best-fit white-light-curve parameters for order 1 in Extended Data Table 2. Furthermore, for the spectroscopic light-curve fits, we only considered the region of order 2 with wavelength <0.85 µm, as the 0.85–1.0-µm range is covered at higher signal-to-noise ratio (SNR) by order 1. All errors on each parameter are representative of 1σ (lower 16th and upper 84th percentiles) of the fit.

chromatic_fitting

chromatic_fitting is an open-source Python tool for modelling multi-wavelength photometric light curves. This tool is built on the framework of chromatic, a package for importing, visualizing and comparing spectroscopic datasets from a variety of sources, including Eureka! and the jwst pipeline. In this paper, we applied chromatic_fitting to the nirHiss reduction.

chromatic_fitting uses the PyMC3 (NUTS) sampler[58] to fit the exoplanet transit model to the light curves. First, we fit the white-light curves for order 1. The white-light curve was generated using an inverse variance-weighted average of the unbinned data. We fixed the orbital period to 4.05528 days (ref. [1]) and assumed a circular orbit. We fit for the mid-transit epoch t_0, the stellar mass M_* and radius R_*, the impact parameter b, the planet-to-star radius ratio R_p/R_*, quadratic limb-darkening coefficients (u_1, u_2) and out-of-transit flux F_0. For the fitted parameters t_0, M_*, R_*, R_p/R_* and F_0, we assumed normal priors N(2459787.56, 0.02^2), N(0.934, 0.056^2), N(0.932, 0.014^2), N(0.146, 0.05^2) and N(1.0, 0.01^2), respectively. For b, we used a uniform prior between 0 and 1.146, in which $b \leq 1 + R_p/R_*$. For the limb-darkening coefficients, we calculated the theoretical values from 3D models in ExoTIC-LD[59,60,61] (based on the stellar parameters $T_{eff} = 5{,}512$ K, log$g$ = 4.47 dex and Fe/H = 0.0 dex (ref. [17])) and assumed normal priors around these values with $\sigma = 0.05$. We also included a second-order polynomial in time to describe the systematics with a fixed constant term of 0.0 and normal priors on the first-order and second-order coefficients c_1 and c_2 of N(0.0, $(1e^{-4})^2$). Using the NUTS implementation of PyMC3, we ran 4,000 tuning

steps and 4,000 draws, with four walkers, for the white-light curve and the mean parameter values are shown in Extended Data Table 2. We checked for convergence using the rank-normalized R-hat diagnostic[62,63].

For each spectroscopic light curve, we fixed the period P, transit epoch t_0, eccentricity e, semi-major axis in stellar radii a/R_* and impact parameter b to the white-light parameter values from the transitspectroscopy/juliet routine (Extended Data Table 2). We then fit for the planet-to-star radius ratio R_p/R_*, quadratic limb-darkening coefficients (u_1, u_2) and out-of-transit flux F_0—for all of these parameters, we assumed the same normal priors as for the white-light curve. We also included a second-order polynomial in time with the same priors as the white-light-curve fit. For each wavelength, we ran 2,000 tuning steps and 2,000 draws, also with four walkers. The final transmission spectrum was taken as the mean value drawn from the posterior distribution for the planet-to-star radius ratio with 1σ uncertainties extracted from the 16th to 84th highest density interval region.

The SNR)in the spectrophotometric light curves from nirHiss for order 1 at 1.34 μm is 165 and that for order 2 at 0.71 μm is 103. We define the SNR as \(\sqrt{{n}_{{\rm{bins}}}}\times {\sigma }_{{\rm{intransit}}}/{\sigma }_{{\rm{outoftransit}}}\), in which σ is the standard deviation.

juliet

We applied the juliet package[64] for light-curve fitting to the products of several reduction pipelines described above. Here we give a general overview of the methods and include exact details for each fit when appropriate.

For the supreme-SPOON reduced stellar spectra, we fit the white-light curve for the mid-transit time, t_0, the impact parameter, b, the scaled orbital semi-major axis, a/R_* and the scaled planetary radius, R_p/R_*, assuming a circular orbit. We also fit two parameters of a quadratic limb-darkening model following the parameterization of ref. [65], as well as an additive scalar jitter and the two parameters of a linear trend with time. We therefore fit seven parameters to the white-light curve for each order, using wide, flat priors for

each case. We then proceeded to fit the light curves from each individual wavelength bin at the native detector resolution—that is, two pixels per bin to roughly account for the extent of the point spread function in the spectral direction. This results in 1,020 bins for order 1 and 520 bins for order 2, as we only consider wavelengths <0.85 µm. For the spectroscopic fits, we fixed the central transit time to the white-light-curve value, and the other orbital parameters were as described for chromatic_fitting. For the linear trend with time, we used the white-light posterior for each of the two parameters as prior distributions for all wavelength bins, whereas for the limb-darkening parameters, we adopted a Gaussian prior centred around the predictions of the ExoTiC-LD package[61,66] with a width of 0.1. As the SOSS throughput files included with ExoTiC-LD did not cover the full wavelength range of both orders, we instead used the throughputs determined during commissioning and included in the spectrace reference file of the jwst pipeline. We truncated the Gaussian prior at 0 and 1, to prevent the limb-darkening parameters from straying into unphysical regions of the parameter space. We then used flat, uninformative priors for the remaining two parameters, the scaled planetary radius and the scalar jitter. The supreme-SPOON white-light-curve fits have $\chi_{\nu}^{2}=1.15$ for order 1 and $\chi_{\nu}^{2}=1.11$ for order 2.

For the transitspectroscopy reduced stellar spectra, we first fit the white-light curves of orders 1 and 2 separately. For these, as suggested above, the period was fixed but the other parameters were allowed to vary. In particular, we set a normal prior on the time-of-transit centre of $N(2459787.5, 0.2^2)$ days, in which the first value denotes the mean and the second the variance of the prior. A normal prior was also set on $a/R_* \sim N(11.37, 0.5^2)$, in which '~' denotes 'distributed as', and a truncated normal between 0 and 1 was set for the impact parameter $b \sim \mathrm{TN}(0.447, 0.1^2)$, in which the means were set following the work of ref. [67] but the variances are large to account for the variation of these parameters in the literature between different authors. We set a uniform prior for the planet-to-star radius ratio between 0 and 0.2 and fixed eccentricity to 0. As well as those, we fit for a mean out-of-transit offset with a normal prior of $N(0, 0.1^2)$ and a jitter term added in quadrature to the error bars with a log-uniform prior between 10 and 1,000 ppm. To account for systematic trends in the data, we use a Gaussian process by means of celerite[68] with a simple Matérn 3/2 kernel to parameterize those

trends. We set log-uniform priors for both the amplitude of this Gaussian process from 10^{-5} to 1,000 ppm and for the timescale from 10^{-3} days to 0.5 days. We use the framework of ref. [65] to parameterize limb darkening through a square-root law, which, following ref. [69], is one of the laws that should give the best results at this level of precision.

For the wavelength-dependent light curves, we used a similar setup with two main differences. The first is that we fix the time-of-transit centre, a/R_* and b to their white-light values. The second is that we set truncated normal priors on the transformed limb-darkening coefficients (q_1, q_2) between 0 and 1, with standard deviations of 0.1 and means obtained by the following method. First, we obtain the nonlinear limb-darkening coefficients using an ATLAS stellar model with the closest properties to those of WASP-39 using the limb-darkening software[70]. Then, the square-root law limb-darkening coefficients are obtained following the algorithm of ref. [71], which are transformed to the (q_1, q_2) parameterization using the equations in ref. [65]. These are then set as the mean for each wavelength-dependent light curve. We note that we fit the light curves at the pixel level, which means fitting one light curve per detector column. We fit them in parallel using the transitspectroscopy.transitfitting.fit_lightcurves routine.

ExoTEP

For the NAMELESS reduction, we perform light-curve fitting on the extracted spectrophotometric observations using the ExoTEP framework[72]. We first fit the white-light curves of both orders 1 and 2 separately. We fit for the mid-transit time t_0, the planet-to-star radius ratio R_p/R_* and quadratic limb-darkening coefficients (u_1, u_2)[73,74], while fixing the impact parameter b and semi-major axis a/R_* to the values of the best order 1 white-light-curve fit from the transitspectroscopy/juliet analysis. We also fit for the scatter σ, as well as a linear systematics model with an offset c and slope v. Uniform priors are considered for all parameters. Furthermore, we only discard the first 10 min of observations (ten integrations) to remove the exponential ramp. For all light curves, we compute the rolling median for a window size of 11 integrations and bring any data point that is more than four standard deviations away from it to the median value. We fit the light curves using the

Markov chain Monte Carlo (MCMC) ensemble sampler emcee[75] for 1,000 steps using four walkers per free parameter. The first 600 steps, 60% of the total amount, are discarded as burn-in. We then fit the spectroscopic light curves, keeping t_0 fixed to its white-light value, at a resolution of three pixels per bin for order 1 (680 bins) and one pixel per bin for order 2 from about 0.6–0.9 μm (675 bins). We used 1,000 steps for the spectroscopic fits, once again discarding the first 600 as burn-in.

iraclis

We analysed all the light curves using the open-source Python package PyLightcurve[76]. For every light curve, PyLightcurve: (1) calculates the limb-darkening coefficients using the ExoTETHyS package[77,78], the wavelength range of the bin, the response curves for each of the NIRISS orders (jwst_niriss_spectrace_0023.fits file, provided by the JWST CRDS) and the stellar parameters ($T_{\mathrm{eff}} = 5{,}540$ K, $\log g = 4.42$ cm s^{-2}, Fe/H = 0.14 dex (ref. [79])); (2) finds the maximum-likelihood model for the data (an exposure-integrated transit model together with a quadratic trend model using the Nelder–Mead minimization algorithm included in the SciPy package[80]; (3) removes outliers that deviate from the maximum-likelihood model by more than three times the standard deviation of the normalized residuals; (4) scales the uncertainties by the RMS of the normalized residuals, to take into account any extra scatter; (5) and, finally, performs an MCMC optimization process using the emcee package[75]. We initially modelled the first-order white-light curve (sum of all bins above 0.85 μm with out-of-transit fluxes above 20 data numbers per second (DN s^{-1}) and fit for the white R_p/R_*, the orbital parameters, a/R_* and i, and the transit mid-time. We then modelled the spectral light curves, fitting only for the R_p/R_*, fixing the orbital parameters, a/R_* and i, and the transit mid-time to the above white results. In both cases, the models also included a quadratic detrending function that was multiplied by the transit model. After modelling, we applied a spectral decontamination step, taking advantage of the varying total flux across the spectral traces. Owing to the contamination, we have $(R_p/R_*)^2 \times (\mathrm{TF} - x)/\mathrm{TF}$, in which TF is the out-of-transit flux (star and contamination) and x is the flux of the contaminating source. Hence, for each wavelength, we fitted for x

and applied the correction ${({R}_{{\rm{p}}}/{R}_{* })}_{{\rm{corr}}}^{2}={({R}_{{\rm{p}}}/{R}_{* })}^{2}\times {\rm{TF}}/({\rm{TF}}-x)$. This procedure is effective in removing uniform contamination. The uniform contamination fixes issues of sky background overcorrection or undercorrection. It also corrects for order overlap. After the decontamination described above, there was still a contaminating source affecting the spectrum around 0.72 μm, which was not uniform because of the point spread function. To separate this source, we applied independent component analysis on the stellar spectra extracted from various distances from the trace. We used two components to describe the contaminating source and one to describe the stellar spectrum. Finally, we estimated the $(R_p/R_*)^2$ for each wavelength bin using the weighted average of all the bins that had the same wavelength range. We only took into account the bins that had out-of-transit fluxes above 20 DN s^{-1}. This choice effectively applied an optimal aperture size for each wavelength bin.

FIREFLy

To extract the transmission spectrum, we bin the cleaned spectrophotometric light curves by wavelength first to create 120 variable-width spectral channels with roughly equal counts in each. We fit for the transit depth of the planet in each channel using a joint light curve and systematics model. The systematics model accounts for spectral shifts in the *X* and *Y* directions[57]. We use the orbital parameters recovered from an MCMC fit to the white-light curve and fix them at each wavelength channel for our fit. We fit for the two quadratic limb-darkening terms *a* and *b* at each wavelength channel. We find that the best-fit limb-darkening coefficients are uniquely determined and deviate by a constant offset relative to model coefficients. Our fits are performed iteratively using the Python package lmfit. The light curves show a typical photometric scatter of 0.3% per integration and the typical transit-depth uncertainties vary between 150 and 300 ppm, which is in line with near-photon-limited precision. More details of the FIREFLy fitting routine can be found in ref. [57] and in ref. [23].

Atmospheric models

To interpret the measured transmission spectrum, we performed an extensive comparison with grids of synthetic transmission spectra. We tested several independent atmospheric models to avoid any model-dependent interpretation of the data. Unless otherwise noted, all of our grids have assumed radiative–convective–thermochemical equilibrium to estimate atmospheric compositions. The exploration of atmospheric models with fewer assumptions (for example, without the assumption of chemical equilibrium with metallicity and C/O ratio as free parameters) and those considering other effects of disequilibrium chemistry is left for future work.

We derive basic interpretations for the observed spectrum based on four independent model grids, ATMO, PHOENIX, PICASO and ScCHIMERA. Each grid contains precomputed transmission spectra at various atmospheric properties, such as M/H, C/O ratio and cloud properties, using from grey to Mie-scattering cloud opacity (see next subsection for details). The ScCHIMERA grid considers further model advancements: (1) various cloud treatments, including grey cloud, grey + power-law cloud opacity and physically motivated (that is, droplet sedimentation) cloud model, (2) the impact of inhomogeneous cloud coverage along the planetary terminator and (3) K/O ratio as a grid dimension. ScCHIMERA provides the best fit to the observations compared with the other three grids and informs the results presented in the main text.

Grid search with precomputed forward models

Here we introduce the independent grids of precomputed transmission spectra, their model description and the main results from these grid fits. We first present the three grids that assume horizontally homogeneous clouds.

ATMO

The atmospheric pressure–temperature (PT) profile is computed using the 1D radiative–convective equilibrium model ATMO[81,82,83]. The model includes the molecular/atomic opacity of CH_4, CO, CO_2, C_2H_2, Cs, FeH, HCN, H_2O, H_2S, K, Li, Na, NH_3, PH_3, Rb, SO_2, TiO and VO, for which the adopted line list is summarized in ref. [84]. The line lists of several key species are: H_2O (ref. [85]), CH_4 (ref. [86]), CO_2 (ref. [87]), CO from HITEMP2010 (ref.

[88]) and K from VALD3 (ref. [89]). We considered atmospheric metallicities M/H = −1.0, +0.0, +1.0, +1.7, +2.0 and +2.3, C/O ratios 0.35, 0.55, 0.7, 0.75, 1.0 and 1.5, planetary intrinsic temperatures T_{int} = 100, 200, 300 and 400 K and day–night energy redistribution factors of 0.25, 0.50, 0.75 and 1.00, for which full heat redistribution corresponds to 0.5. The model varies the O/H ratio to achieve each C/O ratio with a fixed C/H ratio, which is fixed then scaled to solar metallicity. The cloudy models include small particle opacity as the Rayleigh scattering gas opacity enhanced by a factor of either 0 or 10, whereas large particle opacity is equated to the H_2 Rayleigh scattering opacity at 0.35 µm enhanced by a factor of 0.5, 1.0, 5.0, 10.0, 30.0 and 50.0. In total, the ATMO grid consists of 484 cloud-free and 6,292 cloudy atmosphere models. We only consider horizontally homogeneous clouds in the ATMO grid fits.

PHOENIX

The atmospheric PT profile is computed using the 1D radiative–convective equilibrium model PHOENIX[90,91,92]. We considered atmospheric metallicities of M/H = −1.0, +0.0, +1.0 and +2.0, C/O ratio ranging from 0.3 to 1.0 divided into 136 grid points, planetary intrinsic temperatures T_{int} = 200 and 400 K and day–night energy redistribution factors of 0.172, 0.25 and 0.351, in which full heat redistribution corresponds to 0.25. The model varies the C/H ratio to achieve each C/O ratio with a fixed O/H ratio, which is also scaled to solar metallicity. The model includes various chemical species: CH, CH_4, CN, CO, CO_2, COF, C_2, C_2H_2, C_2H_4, C_2H_6, CaH, CrH, FeH, HCN, HCl, HF, HI, HDO, HO_2, H_2, H_2S, H_2O, H_2O_2, H_3^+, MgH, NH, NH_3, NO, N_2, N_2O, OH, O_2, O_3, PH_3, SF_6, SiH, SiO, SiO_2, TiH, TiO, VO and atoms up to U. The line list of H_2O is from BT2 (ref. [85]), other molecular lines from HITRAN 2008 (ref. [93]) and atomic lines from the database of Kurucz and Bell[94]. For cloudy models, the small non-grey cloud particle opacity is treated as a sum of Rayleigh scattering opacity of all gas species enhanced by a factor of either 0 (clear atmosphere) or 10; large grey particle opacity is treated as grey cloud deck pressure levels of 0.3, 3.0 and 10.0 mbar. In total, the PHOENIX grid consists of 95 cloud-free and 380 cloudy atmosphere models. We only consider horizontally homogeneous clouds in the PHOENIX grid fits.

PICASO 3.0

Similarly to the grids of models presented above, we precomputed atmospheric PT profiles using the 1D radiative–convective equilibrium model PICASO 3.0 (refs. [95,96,97,98]) for atmospheric metallicities M/H = −1.0, −0.5, +0.0, +0.5, +1.0, +1.5, +1.7 and +2.0, atmospheric bulk C/O ratios 0.229, 0.458, 0.687 and 1.100, planetary intrinsic temperatures T_{int} = 100, 200 and 300 K and heat redistribution factors of 0.5 and 0.4, in which full heat redistribution corresponds to 0.5. The model fixes the sum of C and O abundances (for example, the (C+O)/H ratio) to that scaled by the metallicity and solar C+O abundance. The model includes 29 chemical species: CH_4, CO, CO_2, C_2H_2, C_2H_4, C_2H_6, CrH, Cs, Fe, FeH, HCN, H_2, H_2O, H_2S, H_3^+, OCS, K, Li, LiCl, LiH, MgH, NH_3, N_2, Na, PH_3, Rb, SiO, TiO and VO. The line lists of several key species are: H_2O (ref. [99]), CH_4 (ref. [100]), CO_2 (ref. [101]), CO (ref. [102]) and K from VALD3 (ref. [89]). For cloudy models, we post-processed the computed PT profiles using the droplet sedimentation model Virga[15,103], which determines the vertical distributions of cloud-mass mixing ratio and mean particle size from the balance between downward mass flux of gravitational sedimentation and upward mass flux of eddy diffusion. We vary vertically constant eddy diffusion coefficients of $K_{zz} = 10^5$, 10^7, 10^9 and 10^{11} and vertically constant sedimentation parameters of f_{sed} = 0.6, 1.0, 3.0, 6.0 and 10.0. The f_{sed} value is defined as the ratio of the mass-averaged sedimentation velocity of cloud particles to the mean upward velocity of the atmosphere, with a smaller f_{sed} yielding more vertically extended clouds[99]; see, for example, refs. [103,104]. We have assumed horizontally homogeneous clouds and accounted for the formation of $MgSiO_3$, MnS and Na_2S clouds. Then, we post-processed the atmospheric properties to compute synthetic transmission spectra. We note that the optical properties of the flux-balanced cloud are computed by the Mie theory[105] under the assumption of a log-normal particle-size distribution with a mean particle size translated from f_{sed} (ref. [15]). In total, the PICASO grid consists of 192 cloud-free and 3,840 cloudy atmosphere models.

We compare the NIRISS/SOSS spectrum (binned to $R = 300$) to each of these model grids and summarize the best fits in the top panel of Extended

Data Fig. 6. For each cloudy and clear model we tested, we compute χ^2/N_{obs} = 2.98–8.55 between the data and the models, with specific values per model indicated in the legend of Extended Data Fig. 6. All of our forward model grids consistently indicate super-solar metallicity (M/H = 1–2) and sub-solar C/O ratio. Each best-fit spectrum shows different structures at >2 μm, as the spectra at these wavelengths are more sensitive to the treatment of cloud properties (see next subsection for details). The best-fit spectra from PICASO, ATMO and PHOENIX indicate atmospheric metallicities of M/H = 1.7, 1.0 and 2.0, respectively. These models also consistently indicate that the C/O ratio is between 0.229 and 0.389, corresponding to the lowest C/O ratio grid point in each grid (see the main text for why models prefer lower C/O ratios). Thus, the super-solar metallicity and sub-solar C/O ratio of WASP-39b are consistent across the different model interpretations of the NIRISS/SOSS transmission spectrum.

We also find that clear atmospheric models fail to fit the observed spectrum even at very high metallicity (M/H = 2.0), as shown in the bottom panel of Extended Data Fig. 6. The clear models fail to match the amplitudes of H_2O absorption features at $\lambda = 0.90$, 1.15, 1.40 and 1.80 μm simultaneously. The clear ATMO models fit the data better than the clear PICASO and PHOENIX models because the ATMO grid allows lower heat redistribution factors (that is, cooler atmosphere). The clear models also overestimate the transit depth at $\lambda \approx 2$ μm because of a strong CO_2 absorption resulting from the inferred high metallicity (M/H = 2.0). The inability of clear atmosphere models to fit the overall NIRISS spectrum strongly indicates the presence of clouds in the atmosphere and emphasizes the ability of the NIRISS wavelength coverage to break the cloud property–metallicity degeneracy. The best-fit cloud properties are $f_{sed} = 1$ and $K_{zz} = 10^9$ cm^2 s^{-1} for Virga clouds in PICASO, a grey cloud opacity of five times the H_2 Rayleigh scattering opacity at 0.35 μm for ATMO and a grey cloud top pressure of 3×10^{-4} bar for PHOENIX.

Grid search with ScCHIMERA

The NIRISS transmission spectrum offers key insights into the atmospheric properties of WASP-39b over a broad wavelength range. The simultaneous

detection of H_2O and K, alongside possible indications of carbon-bearing species, allows us to explore equilibrium models for which the K/O ratio is an extra dimension besides the commonly used C/O ratio and metallicity parameters. Furthermore, as explained in the previous subsection (see also Fig. 3 demonstrating how clouds contribute to the NIRISS spectrum), the broad wavelength coverage of these NIRISS observations makes it possible to explore more complex cloud models beyond traditional grey and homogeneous cloud models. To explore these considerations, we implement the ScCHIMERA grid as explained below.

ScCHIMERA

Previous implementations of this framework include refs. [106,107,108], in which the methods are described in detail. Implementations of this procedure to the JWST data include ref. [109]. For a given set of planetary parameters, our methods precompute the temperature–pressure structure of the planetary atmosphere and the thermochemical equilibrium gas-mixing-ratio profiles. The computations are performed on a grid of atmospheric metallicity (M/H, for example, $\log_{10}$ enrichment relative to solar[20]) spaced at 0.125 dex values between 0 and 2.25 (for example, 1–177 times solar) and C/O ratios at values of 0.20, 0.35, 0.45, 0.55, 0.65, 0.70 and 0.80. Unlike previous implementations of this framework, and to better understand the NIRISS/SOSS observations presented, we include a dimension to our grid exploring the K/O ratio (that is, $\log_{10}$ enrichment relative to solar[20]) with spacing of 0.5 dex between −1 and 0 and 0.1 dex between 0 and 1, overall spanning a range from −1 to 1 or 0.1–10 times solar. In these calculations, the atmospheric metallicity scales the sum of K, C and O. This sum determines the final elemental abundances after scaling metallicity, C/O ratio and K/O ratio. That is, the total oxygen elemental abundance is ${\rm{O}}^{\prime} = \frac{{\rm{M/H}}}{{\rm{K/O\; +\; C/O\; +\; 1}}}$, the total carbon elemental abundance is $C' = O' \times C/O$ and the total potassium elemental abundance is $K' = O' \times K/O$. Furthermore, we explore the energy redistribution (f) between the day and night sides of the planet[110], with values of 0.657, 0.721, 0.791, 0.865, 1.000, 1.030, 1.120, 1.217 and 1.319 in our grid, for which $f = 1.0$ and 2.0 correspond to full day-to-night heat redistribution and dayside-only redistribution, respectively.

The transmission spectrum of the planet is computed with CHIMERA[104,111,112,113,114,115] using the converged atmospheric structures. We compare the observations to these models in a Bayesian inference framework using the nested sampling algorithm MultiNest[108] through its Python implementation PyMultiNest[109] and obtain an optimal set of M/H, C/O ratio, K/O ratio and *f* through nearest-neighbour search in the grid. When computing the transmission spectrum for a given set of (M/H, C/O ratio, K/O ratio, *f*), we also adjust the 1-bar planetary radius controlling the absolute transit depth (an arbitrary pressure with no direct impact on the inferred properties; see, for example, ref. [8]) and model different cloud treatments. The opacity sources considered are H_2–H_2 and H_2–He CIA[116], H_2O (refs. [99,117]), CO_2 (ref. [117]), CO (ref. [88]), CH_4 (ref. [88]), H_2S (ref. [118]), HCN (ref. [119]), Na (refs. [120,121]) and K (refs. [120,122]), which were computed following the methods described in refs. [123,124]. The cloud models considered are: (1) a basic cloud model with a grey, uniformly vertically distributed cloud opacity (κ_{cloud}); (2) a grey + power-law cloud model that accounts for non-grey opacity of small-size particles as a vertically uniform power-law opacity (that is, a parameter for the scattering slope and a Rayleigh enhancement factor; for example, refs. [13,36,39,125]) in addition to grey cloud component, which is expressed by a grey cloud deck of infinite opacity at a given atmospheric pressure; and (3) a droplet sedimentation model[15] (assuming enstatite grains) in which parameters capture the eddy diffusion coefficient and the ratio of sedimentation velocity to characteristic vertical mixing velocity (see also the description of PICASO above). For cloud treatments 2 and 3, we also consider the possibility of inhomogeneities around the planetary limb by considering a linear combination of clear and cloudy models[14], which is key for breaking degeneracies between metallicity and cloud properties[8,40]. We assume the same PT profile for both cloudy and clear limbs in the inhomogeneous cloud models and leave investigation on the possibility of different PT profiles in those regions to future studies.

Identification of absorbers and model selection

We perform our Bayesian inference using all model combinations with the ScCHIMERA grid on four different data resolutions for the nirHiss

transmission spectrum: $R = 100$, $R = 300$, native instrument resolution ($R_{\text{order 1}} = 910$; $R_{\text{order 2}} = 830$) and pixel-level resolution ($R_{\text{order 1}} = 1{,}820$; $R_{\text{order 2}} = 1{,}660$). Resolutions are given at the reference wavelengths of $\lambda = 1.791$ µm for order 1 and 0.744 µm for order 2. We test the robustness of our inferences against different binning and convolution strategies and find the results, that is, the bulk atmospheric properties M/H, C/O ratio and K/O ratio, to be consistent regardless of the resolution of the data. We find a fiducial combination of parameters that can best explain the spectrum (that we call the reference model) with full redistribution ($f = 1$, matching predictions that planets in this temperature regime are unlikely to possess strong day-to-night temperature contrast[126,127,128]), M/H = 1.375 (that is, about 20 times solar), C/O ratio of 0.2 and K/O ratio of 0.1. With these atmospheric properties, the data are best explained by the droplet sedimentation model (ScCHIMERA cloud model 3) and inhomogeneous cover. However, the grey + power-law model (ScCHIMERA cloud model 2) with inhomogeneous cover provides a comparable fit to the data. We compare sets of models by computing their Bayes factor and converting to a 'sigma' detection significance using the prescription in refs. [39,40]. Using $R = 300$ data, the homogeneous droplet sedimentation model (model 3) is preferred over homogeneous grey cloud (model 1) at $\gtrsim 8\sigma$, which strongly indicates the non-grey nature of cloud opacity. Meanwhile, the inhomogeneous droplet sedimentation model is preferred over the homogeneous droplet sedimentation model cloud at 5σ. This is evidence that, for the same model 3, inhomogeneous cloud coverage is preferred. The inhomogeneous droplet sedimentation model is preferred over all the other tested models across all aforementioned resolutions tested.

We explore the contribution of different chemical species to our reference model by performing the Bayesian inference using the inhomogeneous cloud model 3 and artificially disabling the contribution of a selected chemical species, one at a time. By redoing the Bayesian inference, we are able to compare the Bayesian evidence by computing the Bayes factor and converting to a 'sigma' detection significance described above. We detected H_2O at $>30\sigma$, K at 6.8σ and CO at 3.6σ, but no notable detections of Na, CH_4, CO_2, HCN and H_2S. The best-fit metallicity across all models is about 10–30 times solar, the best fit K/O ratio 1–2 times solar and C/O ratio 0.2. Taking the average and standard deviation of the best-fit results for all 20

runs (that is, five models on four data resolutions), we find an average M/H = 19 times solar with a standard deviation of 5 times solar and an average K/O ratio 1.5 times solar with a standard deviation of 0.26 times solar.

Wavelength sensitivity to inferences

We investigate the dependence of the inferred atmospheric properties on the spectral range of the observations by performing the same Bayesian inferences described above on the spectrum blueward of 2 μm (see Extended Data Fig. 7b). This exercise is repeated on all 20 model–data combinations from ScCHIMERA. With the exception of the solar-to-super-solar K/O ratio, inferences about the atmospheric metallicity, C/O ratio and clouds are primarily driven by the shallower transit depth seen in the λ range 2.1–2.3 μm. This wavelength region is that in which the traces of orders 1 and 2 overlap on the detector. To assess the robustness of our results, we explore different data treatments that could affect the final spectrum. First, we find that there are no zeroth-order background contaminants that could be diluting the transit depth in this region. Second, we extract the transmission spectra and fit for dilution between the orders (supreme-SPOON data reduction) and without accounting for the overlap (supreme-SPOON, nirHiss and transitspectroscopy). The evidence for minimal dilution stems from reducing the data through both methods with the same pipeline (supreme-SPOON), which uses the same steps for the entire reduction process along the way, with the exception of fitting and not fitting for dilution. Both techniques yield similar results in the λ range 2.1–2.3 μm. We note that the contamination from order 2 into order 1 was previously shown to be between 8 and 12 ppm (ref. [50]) and is therefore negligible.

We find that, without the data redward of 2 μm, the M/H value is more scattered across models and resolutions with an average metallicity of 61 times solar for the 20 runs and a standard deviation of 28 times solar. On the other hand, the inference on the C/O ratio remains consistently 0.2 across all models and resolutions. Similarly, the K/O ratio remains solar-to-super-solar, with an average of 1.89 times solar and a standard deviation of 0.29 times solar.

These results confirm the necessity for the broad wavelength coverage of NIRISS to constrain the atmospheric metallicity of a planet[8,14,40]. Without the transit depth decrease at 2.1 μm, our models do not exhibit a preference

for cloud models 2 and 3 over cloud model 1, nor do they prefer the presence of inhomogeneities in the cloud cover. Without these constraints on the cloud properties, a wide range of metallicities can provide an equally good fit to the observations blueward of 2 μm when combined with different cloud properties, preventing reliable constraints on the metallicity.

The exploration of these models is summarized in Extended Data Fig. 7. The top panel shows the different cloud treatments and their goodness of fit to the data. Overall, models with inhomogeneous cloud cover best explain the data, with the flux-balanced cloud of model 3 giving the lowest χ^2. The bottom panel contrasts the reference model against the results from all cloud models when using data blueward of 2 μm only. Without the information contained in the dip in transit depth at 2.1 μm, all cloud treatments provide an equally good fit and overestimate the transit depth between 2.0 and 2.3 μm.

K/O ratio inferences

We explore the possibility of constraining the K/O ratio using NIRISS/SOSS. As explained above, across different models and data resolutions, our results indicate that the observations of WASP-39b are best explained by a solar-to-super-solar K/O ratio. To further explore this, we repeat our Bayesian inference for all 20 model–data configurations (five models each at four resolutions) using the observations blueward of 0.8 μm. From high-resolution to low-resolution observations and for all cloud model configurations, we find that all 20 runs prefer models with solar or super-solar K/O ratios for WASP-39b ranging from 1 to 10 times solar. The average across the 20 runs is 2.12 times solar and a standard deviation of 2.33 times solar, with the relatively larger standard deviation resulting from two inferences of highly super-solar K/O ratios of 7 times solar or greater for observations at pixel-level resolution.

Using the reference model atmospheric properties (such as M/H = 1.37, C/O ratio 0.2, full redistribution $f = 1$), we search for the best-fit K/O ratio while simultaneously adjusting the 1-bar radius and the parameters for the inhomogeneous cloud model 3, when only using the observations blueward of 0.8 μm. The best-fit K/O ratio of 0.4 is consistent with the inferences

using all the data and the data blueward of 2.0 μm only. This model is shown in Extended Data Fig. 8 in green. For the best-fit cloud parameters and 1-bar radius, we compute a series of K/O ratios spanning sub-solar and super-solar values. We compute the fit of each model to the data using χ^2 statistics. We then convert the resulting χ^2 value to a *P*-value. These *P*-values allow us to estimate the agreement between each model and the data. Our results find that sub-solar K/O ratios are disfavoured to 2σ, whereas super-solar values $\gtrsim$0.7 are disfavoured to 5σ.

Data availability

The raw data from this study are publicly available at the Space Telescope Science Institute's Mikulski Archive for Space Telescopes (https://archive.stsci.edu/). The data used to create all of the figures in this manuscript are freely available on Zenodo and GitHub (Zenodo link: https://github.com/afeinstein20/wasp39b_niriss_paper). All further data are available on request.

Code availability

The following are open-source pipelines written in Python available either through the Python Package Index (PyPI) or GitHub that were used throughout this work: Eureka! (https://github.com/kevin218/Eureka); nirHiss (https://github.com/afeinstein20/nirhiss); supreme-SPOON (https://github.com/radicamc/supreme-spoon); transitspectroscopy (https://github.com/nespinoza/transitspectroscopy/tree/dev); iraclis (https://github.com/ucl-exoplanets/Iraclis); juliet (https://github.com/nespinoza/juliet); chromatic (https://github.com/zkbt/chromatic); chromatic_fitting (https://github.com/catrionamurray/chromatic_fitting); ExoTiC-LD[13,57] (https://github.com/Exo-TiC/ExoTiC-LD); ExoTETHyS[125] (https://github.com/ucl-exoplanets/ExoTETHyS); PICASO[92,93] (https://github.com/natashabatalha/picaso); Virga[98,99] (https://github.com/natashabatalha/virga); CHIMERA (https://github.com/mrline/CHIMERA); PyMultiNest

(https://github.com/JohannesBuchner/PyMultiNest); MultiNest (https://github.com/JohannesBuchner/MultiNest).

References

1. Fischer, P. D. et al. HST hot-Jupiter transmission spectral survey: clear skies for cool Saturn WASP-39b. *Astrophys. J.* **827**, 19 (2016).

2. Nikolov, N. et al. VLT FORS2 comparative transmission spectroscopy: detection of Na in the atmosphere of WASP-39b from the ground. *Astrophys. J.* **832**, 191 (2016).

3. Wakeford, H. R. et al. The complete transmission spectrum of WASP-39b with a precise water constraint. *Astron. J.* **155**, 29 (2018).

4. Kirk, J. et al. LRG-BEASTS: transmission spectroscopy and retrieval analysis of the highly inflated Saturn-mass planet WASP-39b. *Astron. J.* **158**, 144 (2019).

5. Tsiaras, A. et al. A population study of gaseous exoplanets. *Astron. J.* **155**, 156 (2018).

6. Fisher, C. & Heng, K. Retrieval analysis of 38 WFC3 transmission spectra and resolution of the normalization degeneracy. *Mon. Not. R. Astron. Soc.* **481**, 4698–4727 (2018).

7. Pinhas, A., Madhusudhan, N., Gandhi, S. & MacDonald, R. H_2O abundances and cloud properties in ten hot giant exoplanets. *Mon. Not. R. Astron. Soc.* **482**, 1485–1498 (2019).

8. Welbanks, L. & Madhusudhan, N. On degeneracies in retrievals of exoplanetary transmission spectra. *Astron. J.* **157**, 206 (2019).

9. Min, M., Ormel, C. W., Chubb, K., Helling, C. & Kawashima, Y. The ARCiS framework for exoplanet atmospheres. Modelling philosophy and retrieval. *Astron. Astrophys.* **642**, A28 (2020).

10. Doyon, R. et al. The JWST Fine Guidance Sensor (FGS) and Near-Infrared Imager and Slitless Spectrograph (NIRISS). *Proc. SPIE* **8442**, 84422R (2012).

11. Stevenson, K. B. et al. Transiting exoplanet studies and community targets for JWST's Early Release Science Program. *Publ. Astron. Soc. Pac.* **128**, 094401 (2016).

12. Bean, J. L. et al. The Transiting Exoplanet Community Early Release Science Program for JWST. *Publ. Astron. Soc. Pac.* **130**, 114402 (2018).

13. Lecavelier des Etangs, A., Pont, F., Vidal-Madjar, A. & Sing, D. Rayleigh scattering in the transit spectrum of HD 189733b. *Astron. Astrophys.* **481**, L83–L86 (2008).

14. Line, M. R. & Parmentier, V. The influence of nonuniform cloud cover on transit transmission spectra. *Astrophys. J.* **820**, 78 (2016).

15. Ackerman, A. S. & Marley, M. S. Precipitating condensation clouds in substellar atmospheres. *Astrophys. J.* **556**, 872–884 (2001).

16. Faedi, F. et al. WASP-39b: a highly inflated Saturn-mass planet orbiting a late G-type star. *Astron. Astrophys.* **531**, A40 (2011).

17. Mancini, L. et al. The GAPS programme with HARPS-N at TNG. XVI. Measurement of the Rossiter-McLaughlin effect of transiting planetary systems HAT-P-3, HAT-P-12, HAT-P-22, WASP-39, and WASP-60. *Astron. Astrophys.* **613**, A41 (2018).

18. Biazzo, K. et al. The GAPS Programme with HARPS-N at TNG. XXXV. Fundamental properties of transiting exoplanet host stars. *Astron. Astrophys.* **664**, A161 (2022).

19. Polanski, A. S., Crossfield, I. J. M., Howard, A. W., Isaacson, H. & Rice, M. Chemical abundances for 25 JWST exoplanet host stars with KeckSpec. *Res. Not. Am. Astron. Soc.* **6**, 155 (2022).

20. Lodders, K., Palme, H. & Gail, H.-P. Abundances of the elements in the Solar System. *Landolt Börnstein* **4B**, 712 (2009).

21. Ahrer, E.-M. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRCam. *Nature* https://doi.org/10.1038/s41586-022-05590-4 (2023).

22. Alderson, L. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec G395H. *Nature* https://doi.org/10.1038/s41586-022-05591-3 (2023).

23. Rustamkulov, Z. et al. Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM. *Nature* https://doi.org/10.1038/s41586-022-05677-y (2023).

24. Schneider, A. D. & Bitsch, B. How drifting and evaporating pebbles shape giant planets. II. Volatiles and refractories in atmospheres. *Astron. Astrophys.* **654**, A72 (2021).

25. Hands, T. O. & Helled, R. Super stellar abundances of alkali metals suggest significant migration for hot Jupiters. *Mon. Not. R. Astron. Soc.* **509**, 894–902 (2022).

26. Lothringer, J. D. et al. A new window into planet formation and migration: refractory-to-volatile elemental ratios in ultra-hot Jupiters. *Astrophys. J.* **914**, 12 (2021).

27. Welbanks, L. et al. Mass-metallicity trends in transiting exoplanets from atmospheric abundances of H_2O, Na, and K. *Astrophys. J. Lett.* **887**, L20 (2019).

28. Robertson, P., Bender, C., Mahadevan, S., Roy, A. & Ramsey, L. W. Proxima Centauri as a benchmark for stellar activity indicators in the near-infrared. *Astrophys. J.* **832**, 112 (2016).

29. Changeat, Q. et al. Five key exoplanet questions answered via the analysis of 25 hot-Jupiter atmospheres in eclipse. *Astrophys. J. Suppl. Ser.* **260**, 3 (2022).

30. Lavvas, P., Koskinen, T. & Yelle, R. V. Electron densities and alkali atoms in exoplanet atmospheres. *Astrophys. J.* **796**, 15 (2014).

31. Parmentier, V., Fortney, J. J., Showman, A. P., Morley, C. & Marley, M. S. Transitions in the cloud composition of hot Jupiters. *Astrophys. J.* **828**, 22 (2016).

32. Powell, D. et al. Transit signatures of inhomogeneous clouds on hot Jupiters: insights from microphysical cloud modeling. *Astrophys. J.* **887**, 170 (2019).

33. Roman, M. T. et al. Clouds in three-dimensional models of hot Jupiters over a wide range of temperatures. I. Thermal structures and broadband phase-curve predictions. *Astrophys. J.* **908**, 101 (2021).

34. Demory, B.-O. et al. Inference of inhomogeneous clouds in an exoplanet atmosphere. *Astrophys. J. Lett.* **776**, L25 (2013).

35. Sing, D. K. et al. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature* **529**, 59–62 (2016).

36. MacDonald, R. J. & Madhusudhan, N. HD 209458b in new light: evidence of nitrogen chemistry, patchy clouds and sub-solar water. *Mon. Not. R. Astron. Soc.* **469**, 1979–1996 (2017).

37. Benneke, B. et al. A sub-Neptune exoplanet with a low-metallicity methane-depleted atmosphere and Mie-scattering clouds. *Nat. Astron.* **3**, 813–821 (2019).

38. Barstow, J. K. Unveiling cloudy exoplanets: the influence of cloud model choices on retrieval solutions. *Astrophys. J.* **497**, 4183–4195 (2020).

39. Welbanks, L. & Madhusudhan, N. Aurora: a generalized retrieval framework for exoplanetary transmission spectra. *Astrophys. J.* **913**, 114 (2021).

40. Benneke, B. & Seager, S. Atmospheric retrieval for super-Earths: uniquely constraining the atmospheric composition with transmission spectroscopy. *Astrophys. J.* **753**, 100 (2012).

41. Kreidberg, L. et al. A precise water abundance measurement for the hot Jupiter WASP-43b. *Astrophys. J. Lett.* **793**, L27 (2014).

42. Fortney, J. J. et al. A framework for characterizing the atmospheres of low-mass low-density transiting planets. *Astrophys. J.* **775**, 80 (2013).

43. Madhusudhan, N., Amin, M. A. & Kennedy, G. M. Toward chemical constraints on hot Jupiter migration. *Astrophys. J. Lett.* **794**, L12 (2014).

44. Mordasini, C., van Boekel, R., Mollière, P., Henning, T. H. & Benneke, B. The imprint of exoplanet formation history on observable present-day spectra of hot Jupiters. *Astrophys. J.* **832**, 41 (2016).

45. Shibata, S., Helled, R. & Ikoma, M. The origin of the high metallicity of close-in giant exoplanets. Combined effects of resonant and aerodynamic shepherding. *Astron. Astrophys.* **633**, A33 (2020).

46. Lodders, K. Solar system abundances and condensation temperatures of the elements. *Astrophys. J.* **591**, 1220–1247 (2003).

47. Woitke, P. et al. Equilibrium chemistry down to 100 K. Impact of silicates and phyllosilicates on the carbon to oxygen ratio. *Astron. Astrophys.* **614**, A1 (2018).

48. Öberg, K. I., Murray-Clay, R. & Bergin, E. A. The effects of snowlines on C/O in planetary atmospheres. *Astrophys. J. Lett.* **743**, L16 (2011).

49. Eistrup, C., Walsh, C. & van Dishoeck, E. F. Setting the volatile composition of (exo)planet-building material. Does chemical evolution in disk midplanes matter? *Astron. Astrophys.* **595**, A83 (2016).

50. Darveau-Bernier, A. et al. ATOCA: an algorithm to treat order contamination. Application to the NIRISS SOSS mode. *Publ. Astron.*

Soc. Pac. **134**, 094502 (2022).

51. Bell, T., Ahrer, E.-M., Brande, J., et al. Eureka! an end-to-end pipeline for JWST time-series observations. *J. Open Source Softw.* **7**, 4503 (2022).

52. van Dokkum, P. G. Cosmic-ray rejection by Laplacian edge detection. *Publ. Astron. Soc. Pac.* **113**, 1420–1427 (2001).

53. Craig, M. et al. astropy/ccdproc: v1.3.0.post1. https://doi.org/10.5281/zenodo.1069648 (2017).

54. Radica, M. et al. APPLESOSS: A Producer of ProfiLEs for SOSS. Application to the NIRISS SOSS mode. *Publ. Astron. Soc. Pac.* **134**, 104502 (2022).

55. Espinoza, N. TransitSpectroscopy. https://doi.org/10.5281/zenodo.6960924 (2022).

56. Tsiaras, A. et al. A new approach to analyzing HST spatial scans: the transmission spectrum of HD 209458 b. *Astrophys. J.* **832**, 202 (2016).

57. Rustamkulov, Z., Sing, D. K., Liu, R. & Wang, A. Analysis of a JWST NIRSpec lab time series: characterizing systematics, recovering exoplanet transit spectroscopy, and constraining a noise floor. *Astrophys. J. Lett.* **928**, L7 (2022).

58. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).

59. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for $-5.0 \leq \log[M/H] \leq +1$, 2000 K $\leq T_{eff} \leq$ 50000 K at several surface gravities. *Astron. Astrophys.* **363**, 1081–1190 (2000).

60. Sing, D. K. Stellar limb-darkening coefficients for CoRot and Kepler. *Astron. Astrophys.* **510**, A21 (2010).

61. Laginja, I. & Wakeford, H. ExoTiC-ISM: a Python package for marginalised exoplanet transit parameters across a grid of systematic instrument models. *J. Open Source Softw.* **5**, 2281 (2020).

62. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).

63. Vehtari, A., Gelman, A., Simpson, D., et al. Rank-normalization, folding, and localization: an improved R for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16**, 667–718 (2021).

64. Espinoza, N., Kossakowski, D. & Brahm, R. Juliet: a versatile modelling tool for transiting and non-transiting exoplanetary systems. *Mon. Not. R. Astron. Soc.* **490**, 2262–2283 (2019).

65. Kipping, D. M. Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. *Mon. Not. R. Astron. Soc.* **435**, 2152–2160 (2013).

66. Wakeford, H. & Grant, D. Exo-TiC/ExoTiC-LD: ExoTiC-LD v2.1 Zenodo release. https://doi.org/10.5281/zenodo.6809899 (2022).

67. Maciejewski, G. et al. New transit observations for HAT-P-30 b, HAT-P-37 b, TrES-5 b, WASP-28 b, WASP-36 b, and WASP-39 b. *Acta Astron.* **66**, 55–74 (2016).

68. Foreman-Mackey, D., Agol, E., Ambikasaran, S. & Angus, R. Fast and scalable Gaussian process modeling with applications to astronomical time series. *Astron. J.* **154**, 220 (2017).

69. Espinoza, N. & Jordán, A. Limb darkening and exoplanets – II. Choosing the best law for optimal retrieval of transit parameters. *Mon. Not. R. Astron. Soc.* **457**, 3573–3581 (2016).

70. Espinoza, N. & Jordán, A. Limb darkening and exoplanets: testing stellar model atmospheres and identifying biases in transit parameters. *Mon. Not. R. Astron. Soc.* **450**, 1879–1899 (2015).

71. Howarth, I. D. On stellar limb darkening and exoplanetary transits. *Mon. Not. R. Astron. Soc.* **418**, 1165–1175 (2011).

72. Benneke, B. et al. Water vapor and clouds on the habitable-zone sub-Neptune exoplanet K2-18b. *Astrophys. J. Lett.* **887**, L14 (2019).

73. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).

74. Kreidberg, L. batman: BAsic Transit Model cAlculatioN in Python. *Publ. Astron. Soc. Pac.* **127**, 1161–1165 (2015).

75. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pac.* **125**, 306 (2013).

76. Tsiaras, A. et al. pylightcurve: exoplanet lightcurve model. https://www.ascl.net/1612.018 (2016).

77. Morello, G. et al. ExoTETHyS: tools for exoplanetary transits around host stars. *J. Open Source Softw.* **5**, 1834 (2020).

78. Morello, G. et al. The ExoTETHyS package: tools for exoplanetary transits around host stars. *Astron. J.* **159**, 75 (2020).

79. Hellier, C. et al. Transiting hot Jupiters from WASP-South, Euler and TRAPPIST: WASP-95b to WASP-101b. *Mon. Not. R. Astron. Soc.* **440**, 1982–1992 (2014).

80. Virtanen, P. et al. scipy/scipy: SciPy 1.5.3. https://doi.org/10.5281/zenodo.4100507 (2020).

81. Tremblin, P. et al. Fingering convection and cloudless models for cool brown dwarf atmospheres. *Astrophys. J. Lett.* **804**, L17 (2015).

82. Drummond, B. et al. The effects of consistent chemical kinetics calculations on the pressure-temperature profiles and emission spectra of hot Jupiters. *Astron. Astrophys.* **594**, A69 (2016).

83. Goyal, J. M. et al. A library of ATMO forward model transmission spectra for hot Jupiter exoplanets. *Mon. Not. R. Astron. Soc.* **474**, 5158–5185 (2018).

84. Goyal, J. M. et al. A library of self-consistent simulated exoplanet atmospheres. *Mon. Not. R. Astron. Soc.* **498**, 4680–4704 (2020).

85. Barber, R. J., Tennyson, J., Harris, G. J. & Tolchenov, R. N. A high-accuracy computed water line list. *Mon. Not. R. Astron. Soc.* **368**, 1087–1094 (2006).

86. Yurchenko, S. N. & Tennyson, J. ExoMol line lists – IV. The rotation–vibration spectrum of methane up to 1500 K. *Mon. Not. R. Astron. Soc.* **440**, 1649–1661 (2014).

87. Tashkun, S. A. & Perevalov, V. I. CDSD-4000: high-resolution, high-temperature carbon dioxide spectroscopic databank. *J. Quant. Spectrosc. Radiat. Transf.* **112**, 1403–1410 (2011).

88. Rothman, L. S. et al. HITEMP, the high-temperature molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **111**, 2139–2150 (2010).

89. Ryabchikova, T. et al. A major upgrade of the VALD database. *Phys. Scr.* **90**, 054005 (2015).

90. Hauschildt, P. H., Allard, F. & Baron, E. The NextGen model atmosphere grid for $3000 \leq T_{\mathrm{eff}} \leq 10{,}000$ K. *Astrophys. J.* **512**, 377–385 (1999).

91. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885–895 (2001).

92. Lothringer, J. D. & Barman, T. S. The PHOENIX exoplanet retrieval algorithm and using H^- opacity as a probe in ultrahot Jupiters. *Astron. J.* **159**, 289 (2020).

93. Rothman, L. S. et al. The HITRAN 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **110**, 533–572 (2009).

94. Kurucz, R. & Bell, B. Atomic line data, CD-ROM no. 23. Smithsonian Astrophysical Observatory (1995).

95. McKay, C. P., Pollack, J. B. & Courtin, R. The thermal structure of Titan's atmosphere. *Icarus* **80**, 23–53 (1989).

96. Marley, M. S. & McKay, C. P. Thermal structure of Uranus' atmosphere. *Icarus* **138**, 268–286 (1999).

97. Batalha, N. E., Marley, M. S., Lewis, N. K. & Fortney, J. J. Exoplanet reflected-light spectroscopy with PICASO. *Astrophys. J.* **878**, 70 (2019).

98. Mukherjee, S., Batalha, N. E., Fortney, J. J., et al. PICASO 3.0: a one-dimensional climate model for giant planets and brown dwarfs. *Astrophys. J.* **942**, 71 (2023).

99. Polyansky, O. L. et al. ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *Mon. Not. R. Astron. Soc.* **480**, 2597–2608 (2018).

100. Yurchenko, S. N., Amundsen, D. S., Tennyson, J. & Waldmann, I. P. A hybrid line list for CH_4 and hot methane continuum. *Astron. Astrophys.* **605**, A95 (2017).

101. Huang, X., Gamache, R. R., Freedman, R. S., Schwenke, D. W. & Lee, T. J. Reliable infrared line lists for 13 CO_2 isotopologues up to E'=18,000 cm^{-1} and 1500 K, with line shape parameters. *J. Quant. Spectrosc. Radiat. Transf.* **147**, 134–144 (2014).

102. Li, G. et al. Rovibrational line lists for nine isotopologues of the CO molecule in the $X^1\Sigma^+$ ground electronic state. *Astrophys. J. Suppl. Ser.* **216**, 15 (2015).

103. Rooney, C. M., Batalha, N. E., Gao, P. & Marley, M. S. A new sedimentation model for greater cloud diversity in giant exoplanets and brown dwarfs. *Astrophys. J.* **925**, 33 (2022).

104. Mai, C. & Line, M. R. Exploring exoplanet cloud assumptions in JWST transmission spectra. *Astrophys. J.* **883**, 144 (2019).

105. Bohren, C. F. & Huffman, D. R. *Absorption and Scattering of Light by Small Particles* (Wiley, 1983).

106. Arcangeli, J. et al. H^- opacity and water dissociation in the dayside atmosphere of the very hot gas giant WASP-18b. *Astrophys. J. Lett.* **855**, L30 (2018).

107. Piskorz, D. et al. Ground- and space-based detection of the thermal emission spectrum of the transiting hot Jupiter KELT-2Ab. *Astron. J.* **156**, 133 (2018).

108. Mansfield, M. et al. A unique hot Jupiter spectral sequence with evidence for compositional diversity. *Nat. Astron.* **5**, 1224–1232 (2021).

109. JWST Transiting Exoplanet Community Early Release Science Team Identification of carbon dioxide in an exoplanet atmosphere. *Nature* https://doi.org/10.1038/s41586-022-05269-w (2022).

110. Fortney, J. J. The effect of condensates on the characterization of transiting planet atmospheres with transmission spectroscopy. *Mon. Not. R. Astron. Soc.* **364**, 649–653 (2005).

111. Line, M. R. et al. A systematic retrieval analysis of secondary eclipse spectra. I. A comparison of atmospheric retrieval techniques. *Astrophys. J.* **775**, 137 (2013).

112. Iyer, A. R. & Line, M. R. The influence of stellar contamination on the interpretation of near-infrared transmission spectra of sub-Neptune worlds around M-dwarfs. *Astrophys. J.* **889**, 78 (2020).

113. Feroz, F., Hobson, M. P. & Bridges, M. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* **398**, 1601–1614 (2009).

114. Buchner, J. et al. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astron. Astrophys.* **564**, A125 (2014).

115. Line, M. R. et al. Information content of exoplanetary transit spectra: an initial look. *Astrophys. J.* **749**, 93 (2012).

116. Richard, C. et al. New section of the HITRAN database: collision-induced absorption (CIA). *J. Quant. Spectrosc. Radiat. Transf.* **113**, 1276–1285 (2012).

117. Freedman, R. S. et al. Gaseous mean opacities for giant planet and ultracool dwarf atmospheres over a range of metallicities and temperatures. *Astrophys. J. Suppl. Ser.* **214**, 25 (2014).

118. Azzam Ala'a, A. A., Tennyson, J., Yurchenko, S. N. & Naumenko, O. V. ExoMol molecular line lists – XVI. The rotation–vibration spectrum of hot H_2S. *Mon. Not. R. Astron. Soc.* **460**, 4063–4074 (2016).

119. Barber, R. J. et al. ExoMol line lists - III. An improved hot rotation–vibration line list for HCN and HNC. *Mon. Not. R. Astron. Soc.* **437**, 1828–1835 (2014).

120. Kramida, A., Ralchenko, Y., Reader, J. & NIST ASD Team. *NIST Atomic Spectra Database V 5.6* (National Institute of Standards and Technology, 2018).

121. Allard, N. F., Spiegelman, F., Leininger, T. & Molliere, P. New study of the line profiles of sodium perturbed by H_2. *Astron. Astrophys.* **628**, A120 (2019).

122. Allard, N. F., Spiegelman, F. & Kielkopf, J. F. K–H_2 line shapes for the spectra of cool brown dwarfs. *Astron. Astrophys.* **589**, A21 (2016).

123. Gharib-Nezhad, E. et al. EXOPLINES: molecular absorption cross-section database for brown dwarf and giant exoplanet atmospheres. *Astrophys. J. Suppl. Ser.* **254**, 34 (2021).

124. Grimm, S. L. et al. HELIOS-K 2.0 opacity calculator and open-source opacity database for exoplanetary atmospheres. *Astrophys. J. Suppl. Ser.* **253**, 30 (2021).

125. Ohno, K. & Kawashima, Y. Super-Rayleigh slopes in transmission spectra of exoplanets generated by photochemical haze. *Astrophys. J. Lett.* **895**, L47 (2020).

126. Perez-Becker, D. & Showman, A. P. Atmospheric heat redistribution on hot Jupiters. *Astrophys. J.* **776**, 134 (2013).

127. Komacek, T. D. & Showman, A. P. Atmospheric circulation of hot Jupiters: dayside–nightside Temperature differences. *Astrophys. J.* **821**, 16 (2016).

128. Zhang, X. Atmospheric regimes and trends on exoplanets and brown dwarfs. *Res. Astron. Astrophys.* **20**, 099 (2020).

129. Bradley, L. et al. astropy/photutils: 1.0.0. https://doi.org/10.5281/zenodo.4044744 (2020).

Acknowledgements

This work is based on observations made with the NASA/ESA/CSA JWST. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for the JWST. These observations are associated with programme no. 1366. Support for this programme was provided by NASA through a grant from the Space Telescope Science Institute. A.D.F. acknowledges support by the National Science Foundation Graduate Research Fellowships Program under grant no. DGE-1746045. M.R. acknowledges support from the Natural Sciences and Engineering Research Council of Canada, the

Fonds de Recherche du Québec, as well as the Institut de Recherche sur les Exoplanètes. This research made use of ccdproc, an Astropy package for image reduction[53]. This research made use of Photutils, an Astropy package for the detection and photometry of astronomical sources[129]. This research has made use of NASA's Astrophysics Data System Bibliographic Services.

Author information

Authors and Affiliations

1. Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL, USA

   Adina D. Feinstein & Jacob L. Bean

2. Department of Physics, Université de Montréal, Montreal, Quebec, Canada

   Michael Radica, Louis-Philippe Coulombe, Björn Benneke, Jake Taylor, Caroline Piaulet & Pierre-Alexis Roy

3. Institute for Research on Exoplanets, Université de Montréal, Montreal, Quebec, Canada

   Michael Radica, Louis-Philippe Coulombe, Björn Benneke, Jake Taylor & Caroline Piaulet

4. School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

   Luis Welbanks & Michael R. Line

5. Department of Astrophysical and Planetary Sciences, University of Colorado Boulder, Boulder, CO, USA

   Catriona Anne Murray & Zachory K. Berta-Thompson

6. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

Kazumasa Ohno, Jonathan J. Fortney, Natalie M. Batalha, Aarynn L. Carter & Sagnick Mukherjee

7. Space Telescope Science Institute, Baltimore, MD, USA

Néstor Espinoza, Quentin Changeat, Leonardo A. dos Santos & Nikolay K. Nikolov

8. Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD, USA

Néstor Espinoza & David K. Sing

9. Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA

Johanna K. Teske, Peter Gao & Munazza K. Alam

10. Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA

Zafar Rustamkulov, David K. Sing & Natalie H. Allen

11. Department of Physics and Astronomy, University College London, London, UK

Arianna Saba, Angelos Tsiaras & Quentin Changeat

12. INAF - Osservatorio Astrofisico di Arcetri, Florence, Italy

Angelos Tsiaras

13. School of Physical Sciences, The Open University, Milton Keynes, UK

Joanna K. Barstow

14. Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

Heather A. Knutson

15. Department of Astronomy, University of Michigan, Ann Arbor, MI, USA

Ryan J. MacDonald

16. Department of Astronomy, Cornell University, Ithaca, NY, USA

Ryan J. MacDonald & Jake D. Turner

17. Carl Sagan Institute, Cornell University, Ithaca, NY, USA

Ryan J. MacDonald & Jake D. Turner

18. Max Planck Institute for Astronomy, Heidelberg, Germany

Thomas Mikal-Evans, Laura Kreidberg, Karan Molaverdikhani, Thomas Henning & Luigi Mancini

19. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Benjamin V. Rackham

20. Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Benjamin V. Rackham

21. Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

Jake Taylor, Vivien Parmentier & Xianyu Tan

22. Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Université Côte d'Azur, Nice, France

Vivien Parmentier

23. Astrobiology Program, UC Santa Cruz, Santa Cruz, CA, USA

Natalie M. Batalha

24. European Space Agency, Space Telescope Science Institute, Baltimore, MD, USA

Quentin Changeat & Sarah Kendrew

25. School of Physics, Trinity College Dublin, Dublin, Ireland

Neale P. Gibson

26. School of Earth and Planetary Sciences (SEPS), National Institute of Science Education and Research (NISER), Homi Bhabha National Institute (HBNI), Jatani, India

Jayesh M. Goyal

27. Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA

Mercedes López-Morales, James Kirk & Diana Powell

28. Department of Physics, Utah Valley University, Orem, UT, USA

Joshua D. Lothringer

29. Leiden Observatory, University of Leiden, Leiden, The Netherlands

Yamila Miguel & Nicolas Crouzet

30. SRON Netherlands Institute for Space Research, Leiden, The Netherlands

Yamila Miguel

31. Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Munich, Germany

Karan Molaverdikhani & Kevin Heng

32. Exzellenzcluster Origins, Garching, Germany

Karan Molaverdikhani

33. Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA

Sarah E. Moran

34. Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain

Giuseppe Morello & Enric Palle

35. Departamento de Astrofísica, Universidad de La Laguna (ULL), Tenerife, Spain

Giuseppe Morello

36. INAF - Palermo Astronomical Observatory, Palermo, Italy

Giuseppe Morello

37. Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA

Kevin B. Stevenson & E. M. May

38. School of Physics, University of Bristol, Bristol, UK

Hannah R. Wakeford & Lili Alderson

39. Centre for Exoplanets and Habitability, University of Warwick, Coventry, UK

Eva-Maria Ahrer & Peter J. Wheatley

40. Department of Physics, University of Warwick, Coventry, UK

Eva-Maria Ahrer, Kevin Heng & Peter J. Wheatley

41. NASA Ames Research Center, Moffett Field, CA, USA

Natasha E. Batalha

42. Bay Area Environmental Research Institute, NASA Ames Research Center, Moffett Field, CA, USA

Taylor J. Bell

43. Department of Physics, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Jasmina Blecic

44. Center for Astro, Particle, and Planetary Physics (CAP3), New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Jasmina Blecic

45. Department of Physics & Astronomy, University of Kansas, Lawrence, KS, USA

Jonathan Brande & Ian J. M. Crossfield

46. Instituto de Astrofísica, Universidad Andrés Bello, Santiago, Chile

Claudio Caceres

47. Núcleo Milenio de Formación Planetaria (NPF), Valparaíso, Chile

Claudio Caceres

48. Centro de Astrofísica y Tecnologías Afines (CATA), Santiago, Chile

Claudio Caceres

49. School of Physics and Astronomy, University of Leicester, Leicester, UK

S. L. Casewell & Michael T. Roman

50. Centre for Exoplanet Science, University of St Andrews, St Andrews, UK

Katy L. Chubb

51. INAF - Osservatorio Astrofisico di Torino, Pino Torinese, Italy

Patricio E. Cubillos

52. Space Research Institute, Austrian Academy of Sciences, Graz, Austria

Patricio E. Cubillos

53. Institute of Astronomy, Department of Physics and Astronomy, KU Leuven, Leuven, Belgium

Leen Decin

54. Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, The Netherlands

Jean-Michel Désert

55. Planetary Sciences Group, Department of Physics, University of Central Florida, Orlando, FL, USA

Joseph Harrington

56. Florida Space Institute, University of Central Florida, Orlando, FL, USA

Joseph Harrington

57. ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

Kevin Heng

58. Institute for Astrophysics, University of Vienna, Vienna, Austria

Nicolas Iro

59. Department of Astronomy, University of Maryland, College Park, MD, USA

Eliza M.-R. Kempton & Matthew C. Nixon

60. Department of Physics, Imperial College London, London, UK

James Kirk

61. Infrared Processing and Analysis Center (IPAC), California Institute of Technology, Pasadena, CA, USA

Jessica Krick

62. Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, Gif-sur-Yvette, France

Pierre-Olivier Lagage

63. Département d’Astronomie, Université de Genève Sauverny, Versoix, Switzerland

Monika Lendl & Dominique J. M. Petit dit de la Roche

64. Department of Physics, University of Rome “Tor Vergata”, Rome, Italy

Luigi Mancini

65. INAF - Turin Astrophysical Observatory, Pino Torinese, Italy

Luigi Mancini

66. Steward Observatory, University of Arizona, Tucson, AZ, USA

Megan Mansfield & Everett Schlawin

67. Department of Physics and Astronomy, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK

N. J. Mayne

68. Astronomy Department, Wesleyan University, Middletown, CT, USA

Seth Redfield

69. Van Vleck Observatory, Wesleyan University, Middletown, CT, USA

Seth Redfield

70. Institute of Astronomy, University of Cambridge, Cambridge, UK

Laura K. Rogers & Matthew C. Nixon

71. Universidad Adolfo Ibáñez, Campus Peñalolén, Santiago, Chile

Michael T. Roman

72. Maison de la Simulation, CEA, CNRS, Université Paris-Sud, Université Versailles St Quentin, Université Paris-Saclay, Gif-sur-Yvette, France

P. Tremblin

73. Université de Paris Cité and Université Paris-Est Creteil, CNRS, LISA, Paris, France

Olivia Venot

74. Astrophysics & Planetary Sciences, University of Colorado Boulder, Boulder, CO, USA

William C. Waalkes

75. Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA

Xi Zhang

Contributions

All authors played a substantial role in producing the results of this manuscript. Those roles include: developing the original JWST Early Release Science proposal, management of the project, definition of the target list, designing the observation plan, reduction and analysis of the data, theoretical modelling and preparation of this manuscript. Some specific contributions are listed as follows. N.M.B., J.L.B. and K.B.S. provided overall programme leadership and management. A.D.F., M.R., L.W. and J.L.B. led the efforts for this manuscript. D.K.S., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., K.B.S., L.K., M.L.-M., M.R.L., N.M.B., V.P. and Z.K.B.-T. made notable contributions to the design of the programme. K.B.S. generated the observing plan, with input from the team. N.E. provided instrument expertise. B.B., E.M.-R.K., H.R.W., I.J.M.C., J.L.B., L.K., M.L.-M., M.R.L., N.M.B. and Z.K.B.-T. led or co-led working groups and/or contributed to substantial strategic planning efforts, such as the design and implementation of the prelaunch Data Challenges. A.L.C., D.K.S., E.S., N.E., N.P.G. and V.P. generated simulated data for prelaunch testing of methods. A.D.F., M.R., L.W., C.A.M., K.O., J.L.B. and J.T. contributed substantially to the writing of this manuscript, along with contributions in Methods from L.-P.C., B.B., Z.R., A.S. and A.T. A.D.F., M.R., C.A.M., L.-P.C., N.E., Z.R., A.S. and A.T. contributed to the development of data-analysis pipelines and/or provided the data-analysis

products used in this analysis, that is, reduced the data, modelled the light curves and/or produced the planetary spectrum. L.W., K.O., M.R.L. and S.M. generated theoretical model grids for comparison with data. A.D.F., L.W. and K.O. generated figures for this manuscript. J.B., L.A.D.S., J.J.F., P.G., H.A.K., R.J.M., T.M.-E., B.V.R. and V.P. provided notable feedback to the manuscript, coordinating comments from all other authors.

Corresponding author

Correspondence to Adina D. Feinstein.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Extended data

is available for this paper at https://doi.org/10.1038/s41586-022-05674-1.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Comparison of the (x, y) positions of NIRISS orders 1 and 2 across the detector as modelled from different reduction pipelines.

a,**c**, Each pipeline traces the curvature of orders 1 and 2 using different methods. We show the best-fit trace for order 1 in panel **a** and order 2 in panel **b**. We highlight zoomed-in regions to further examine differences. We note that iraclis uses the JWST-provided spectral trace. There is generally good agreement between the traces across the entire detector (<1 pixel deviations), with the strongest deviations towards the ends of each trace (for example, x pixel position <500 for orders 1 and 2). This demonstrates the reliability of order spectral traces across all pipelines. **b**,**d**, We provide an example spatial profile along the column for order 1 (**b**) and order 2 (**d**) at x = 1,250. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure1.py).

Extended Data Fig. 2 Comparison of averaged background frames computed for each reduction pipeline.

a, A median out-of-transit integration from the stage 2 output files of the jwst pipeline in data numbers per second (DN s^{-1}). **b1**,**c1**,**d1**,**e1**, Estimated median background frames for four example pipelines: nirHiss (**b**), supreme-SPOON (**c**), transitspectroscopy (**d**) and iraclis (**e**). All reduction pipelines use the predefined zodiacal background model provided on the STScI JDox User Documentation website. nirHiss estimates the zeroth-order contaminants by taking a smoothed median from the F277W filter integrations. We note that the background frame from supreme-SPOON (**c1**) for group eight is shown here and was scaled by a factor of about 0.02 to lie on the same scale as the background from the other three pipelines. iraclis subtracts a median per column to remove further $1/f$ noise. **b2**,**c2**,**d2**,**e2**, Background-subtracted median integrations for each pipeline, plotted in DN s^{-1} but scaled from −5 to 5 to highlight subtle changes in the background. For these integrations, we define out of transit as integrations 1–200 and 400–518.

(https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure2.py).

Extended Data Fig. 3 Example extracted stellar spectra from different reduction pipelines.

The insets highlight the peak of the spectra. The supreme-SPOON spectra are scaled by a factor of 72 to compare the overall shape of the spectra, rather than the extracted flux counts. **a**, The extracted stellar spectra from order 1. All pipelines are in relatively good agreement, whereas the shape of the iraclis data changes slightly at $\lambda < 1.1$ µm. This is probably owing to different traces that were used in the spectral extraction routine. **b**, The extracted stellar spectra from order 2. Differences at $\lambda = 0.725$ and 0.860 µm are because of differences in removing zeroth-order contaminants in the background. The iraclis pipeline does not extract data past $\lambda > 0.85$ µm, which is where the order overlap region begins. Across all pipelines, the shape of the spectra, as well as overall absorption features, cosmic-ray-removal techniques and noise levels are consistent. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure3.py).

Extended Data Fig. 4 White-light curves and residuals between the light curve and best-fit exoplanet transit model for each reduction pipeline.

a, White-light curves for order 1 (top) and order 2 (bottom) with the out-of-transit scatter noted in the figure text. **b**, The residuals, in ppm, between the plotted light curves and best-fit exoplanet transit model. The start of transit ingress and end of transit egress are marked with dashed vertical lines; the transit midpoint is marked with a solid vertical line. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure4.py).

Extended Data Fig. 5 Transmission spectra for WASP-39b for all reduction techniques.

Our best-fit reference model to the nirHiss spectrum (red) is plotted as a solid black line in all panels and the spectra are separated into three panels for ease of reading. Wavelengths that overlap with zeroth-order contaminants are masked. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure5.py).

Extended Data Fig. 6 A summary of precomputed forward model fits to the NIRISS/SOSS spectrum.

a, Each coloured line shows the best-fit spectrum from the PICASO, ATMO and PHOENIX cloudy grid. The χ^2 values per number of data points (N_{obs} = 327) are $\chi^2/N_{obs} = 2.98$, 3.24 and 3.51 for the PICASO, ATMO and PHOENIX grids, respectively. All grid models consistently indicate a super-solar metallicity of M/H = 1–2 and a sub-solar C/O ratio. **b**, The same as the top panel but for the best-fit clear atmosphere models. The clear models yield noticeably worse fits to the data, $\chi^2/N_{obs} = 7.02$, 4.11 and 8.55 for the PICASO, ATMO and PHOENIX grids, respectively, which strongly indicates the presence of clouds in the atmosphere. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure6.py).

Extended Data Fig. 7 A demonstration of how the redder wavelength coverage of NIRISS/SOSS drives the inference on cloud structure for WASP-39b.

We fit the NIRISS/SOSS spectrum (grey) using a suite of cloud models to derive the best-fit C/O ratio and metallicity. Here we demonstrate how the best-fit model for each cloud treatment changes as a function of what wavelength region we fit. **a**, The best-fit models when using the entire wavelength coverage of NIRISS/SOSS. **b**, The best-fit models when using $\lambda < 2$ µm, which excludes the overlapping region between orders on the detector. The reference spectrum (black) on both panels corresponds to the best-fit inhomogeneous droplet sedimentation model for the entire wavelength coverage. The fitted data are presented as dark-grey points. The

quoted numbers in brackets in the legend are the respective χ^2/N for each fit for the top (left value) and bottom (right value). The difference between cloud models is within the noise of the NIRISS/SOSS data when fitting to $\lambda < 2$ μm. It is clear that fitting the entire NIRISS/SOSS wavelength coverage results in a lower χ^2/N and better fit. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure7.py).

Extended Data Fig. 8 Evidence for super-solar K/O ratio in WASP-39b.

We fit for the K/O ratio while keeping the rest of the model parameters (for example, C/O ratio, metallicity and redistribution) the same as our reference model and fitting for the cloud parameters and scaled planetary radius. Here we present the different K/O ratio models (solid lines) we fit against the transmission spectrum at $R = 300$ (black and white points). We represent the respective fit of each model in the orange shading. (https://github.com/afeinstein20/wasp39b_niriss_paper/blob/main/scripts/edfigure8.py).

Extended Data Table 1 An outline of reduction and fitting pipelines used to produce transmission spectra for WASP-39b with NIRISS/SOSS

Extended Data Table 2 White-light curve best-fit orbital parameters from order 1

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the

article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Feinstein, A.D., Radica, M., Welbanks, L. *et al.* Early Release Science of the exoplanet WASP-39b with JWST NIRISS. *Nature* **614**, 670–675 (2023). https://doi.org/10.1038/s41586-022-05674-1

- Received: 02 November 2022
- Accepted: 20 December 2022
- Published: 09 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05674-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Early Release Science of the exoplanet WASP-39b with JWST NIRCam**

    - Eva-Maria Ahrer
    - Kevin B. Stevenson
    - Xi Zhang

  Nature (2023)

JWST opens a window on exoplanet skies

- Julia V. Seidel
- Louise D. Nielsen
- Subhajit Sarkar

Nature News & Views Forum 15 Feb 2023

A population of red candidate massive galaxies ~600 Myr after the Big Bang

- Ivo Labbé
- Pieter van Dokkum
- Bingjie Wang

Nature Article 22 Feb 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05674-1

- Article
- Open Access
- Published: 22 February 2023

Suppressing quantum errors by scaling a surface code logical qubit

- Google Quantum AI

Nature **volume 614**, pages 676–681 (2023)

- 32k Accesses

- 629 Altmetric

- Metrics details

Abstract

Practical quantum computing will require error rates well below those achievable with physical qubits. Quantum error correction[1,2] offers a path to algorithmically relevant error rates by encoding logical qubits within many physical qubits, for which increasing the number of physical qubits enhances protection against physical errors. However, introducing more qubits also increases the number of error sources, so the density of errors must be sufficiently low for logical performance to improve with increasing code size. Here we report the measurement of logical qubit performance scaling across several code sizes, and demonstrate that our system of superconducting qubits has sufficient performance to overcome the additional errors from increasing qubit number. We find that our distance-5 surface code logical qubit modestly outperforms an ensemble of distance-3 logical qubits on average, in terms of both logical error probability over 25 cycles and logical error per cycle ((2.914 ± 0.016)% compared to (3.028 ±

0.023)%). To investigate damaging, low-probability error sources, we run a distance-25 repetition code and observe a 1.7×10^{-6} logical error per cycle floor set by a single high-energy event (1.6×10^{-7} excluding this event). We accurately model our experiment, extracting error budgets that highlight the biggest challenges for future systems. These results mark an experimental demonstration in which quantum error correction begins to improve performance with increasing qubit number, illuminating the path to reaching the logical error rates required for computation.

Main

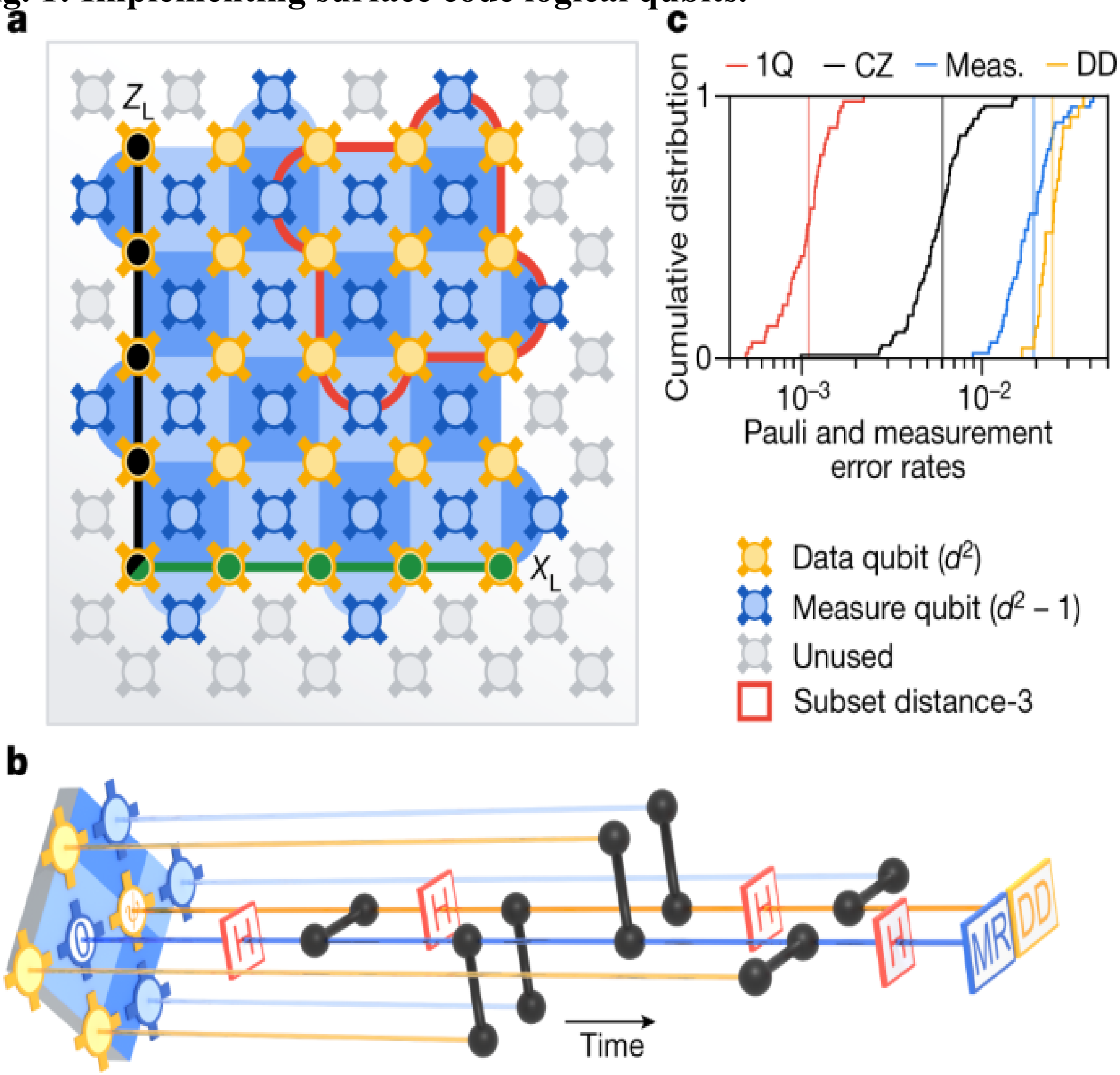
Since Feynman's proposal to compute using quantum mechanics[3], many potential applications have emerged, including factoring[4], optimization[5], machine learning[6], quantum simulation[7] and quantum chemistry[8]. These applications often require billions of quantum operations[9,10,11] and state-of-the-art quantum processors typically have error rates around 10^{-3} per gate[12,13,14,15,16,17], far too high to execute such large circuits. Fortunately, quantum error correction can exponentially suppress the operational error rates in a quantum processor, at the expense of temporal and qubit overhead[18,19].

Several works have reported quantum error correction on codes able to correct a single error, including the distance-3 Bacon–Shor[20], colour[21], five-qubit[22], heavy-hexagon[23] and surface[24,25] codes, as well as continuous variable codes[26,27,28,29]. However, a crucial question remains of whether scaling up the error-correcting code size will reduce logical error rates in a real device. In theory, logical errors should be reduced if physical errors are sufficiently sparse in the quantum processor. In practice, demonstrating reduced logical error requires scaling up a device to support a code that can correct at least two errors, without sacrificing state-of-the-art performance. In this work we report a 72-qubit superconducting device supporting a 49-qubit distance-5 ($d = 5$) surface code that narrowly outperforms its average subset 17-qubit distance-3 surface code, demonstrating a critical step towards scalable quantum error correction.

Surface codes with superconducting qubits

Surface codes[30,31,32,33,34] are a family of quantum error-correcting codes that encode a logical qubit into the joint entangled state of a $d \times d$ square of physical qubits, referred to as data qubits. The logical qubit states are defined by a pair of anti-commuting logical observables X_L and Z_L. For the example shown in Fig. 1a, a Z_L observable is encoded in the joint Z-basis parity of a line of qubits that traverses the lattice from top to bottom, and likewise an X_L observable is encoded in the joint X-basis parity traversing left to right. This non-local encoding of information protects the logical qubit from local physical errors, provided we can detect and correct them.

Fig. 1: Implementing surface code logical qubits.



a, Schematic of a 72-qubit Sycamore device with a distance-5 surface code embedded, consisting of 25 data qubits (gold) and 24 measure qubits (blue). Each measure qubit is associated with a stabilizer (blue coloured tile, dark: X, light: Z). Representative logical operators Z_L (black) and X_L (green) traverse the array, intersecting at the lower-left data qubit. The upper right quadrant (red outline) is one of four subset distance-3 codes (the four quadrants) that we compare to distance-5. **b**, Illustration of a stabilizer measurement, focusing on one data qubit (labelled ψ) and one measure qubit (labelled 0), in perspective view with time progressing to the right. Each qubit participates in four CZ gates (black) with its four nearest neighbours, interspersed with Hadamard gates (H), and finally, the measure qubit is measured and reset to \(\left|0\right\rangle \) (MR). Data qubits perform dynamical decoupling (DD) while waiting for the measurement and reset. All stabilizers are measured in this manner concurrently. Cycle duration is 921 ns, including 25-ns single-qubit gates, 34-ns two-qubit gates, 500-ns measurement and 160-ns reset (see Supplementary Information for compilation details). The readout and reset take up most of the cycle time, so the concurrent data qubit idling is a dominant source of error. **c**, Cumulative distributions of errors for single-qubit gates (1Q), CZ gates, measurement (Meas.) and data qubit dynamical decoupling (idle during measurement and reset), which we refer to as component errors. The circuits were benchmarked in simultaneous operation using random circuit techniques, on the 49 qubits used in distance-5 and the 4 CZ layers from the stabilizer circuit[38,59] (see Supplementary Information). Vertical lines are means.

To detect errors, we periodically measure X and Z parities of adjacent clusters of data qubits with the aid of $d^2 - 1$ measure qubits interspersed throughout the lattice. As shown in Fig. 1b, each measure qubit interacts with its neighbouring data qubits to map the joint data qubit parity onto the measure qubit state, which is then measured. Each parity measurement, or stabilizer, commutes with the logical observables of the encoded qubit as well as every other stabilizer. Consequently, we can detect errors when parity measurements change unexpectedly, without disturbing the logical qubit state.

A decoder uses the history of stabilizer measurement outcomes to infer likely configurations of physical errors on the device. We can then determine

the overall effect of these inferred errors on the logical qubit, thus preserving the logical state. Most surface code logical gates can be implemented by maintaining logical memory and executing different sequences of measurements on the code boundary[35,36,37]. Thus, we focus on preserving logical memory, the core technical challenge in operating the surface code.

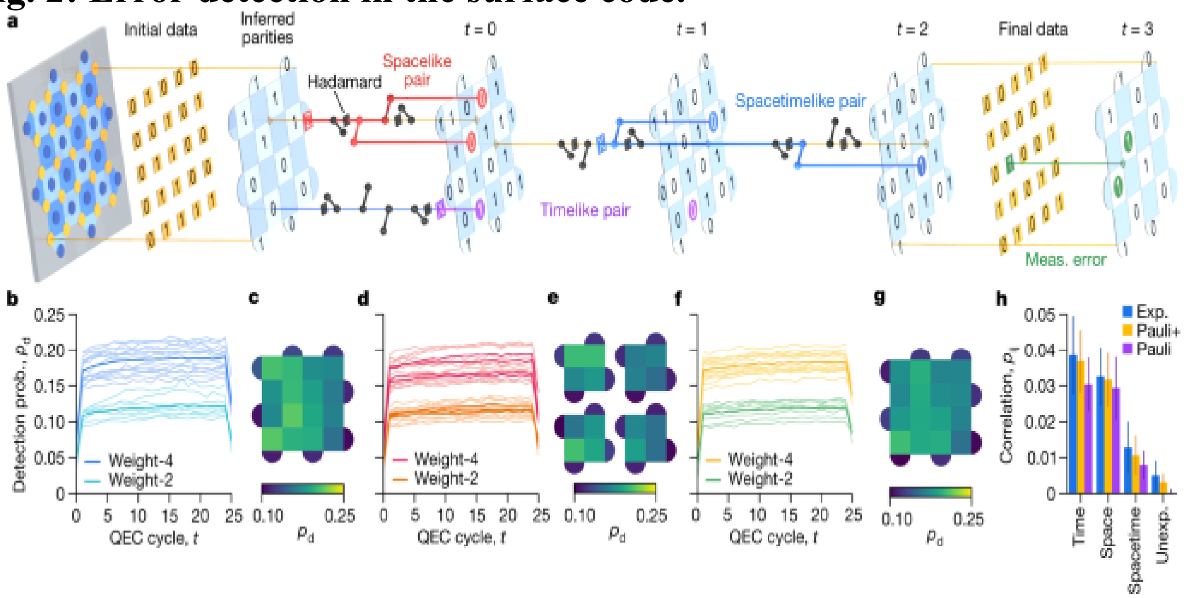
We implement the surface code on an expanded Sycamore device[38] with 72 transmon qubits[39] and 121 tunable couplers[40,41]. Each qubit is coupled to four nearest neighbours except on the boundaries, with mean qubit coherence times $T_1 = 20$ μs and $T_{2,\mathrm{CPMG}} = 30$ μs, in which CPMG represents Carr–Purcell–Meiboom–Gill. As in ref. [42], we implement single-qubit rotations, controlled-Z (CZ) gates, reset and measurement, demonstrating similar or improved simultaneous performance as shown in Fig. 1c.

The distance-5 surface code logical qubit is encoded on a 49-qubit subset of the device, with 25 data qubits and 24 measure qubits. Each measure qubit corresponds to one stabilizer, classified by its basis (X or Z) and the number of data qubits involved (weight, 2 or 4). Ideally, to assess how logical performance scales with code size, we would compare distance-5 and distance-3 logical qubits under identical noise. Although device inhomogeneity makes this comparison difficult, we can compare the distance-5 logical qubit to the average of four distance-3 logical qubit subgrids, each containing nine data qubits and eight measure qubits. These distance-3 logical qubits cover the four quadrants of the distance-5 code with minimal qubit overlap, capturing the average performance of the full distance-5 grid.

In a single instance of the experiment, we initialize the logical qubit state, run several cycles of error correction, and then measure the final logical state. We show an example in Fig. 2a. To prepare a Z_L eigenstate, we first prepare each data qubit in \(\left|0\right\rangle \) or \(\left|1\right\rangle \), an eigenstate of the Z stabilizers. The first cycle of stabilizer measurements then projects the data qubits into an entangled state that is also an eigenstate of the X stabilizers. Each cycle contains CZ and Hadamard gates sequenced to extract X and Z stabilizers simultaneously, and ends with the measurement and reset of the measure qubits. In the final cycle, we also measure the data qubits in the Z basis, yielding both parity information and a measurement of

the logical state. Preparing and measuring X_L eigenstates proceeds analogously. The instance succeeds if the corrected logical measurement agrees with the known initial state; otherwise, a logical error has occurred.

Fig. 2: Error detection in the surface code.



a, Illustration of a surface code experiment, in perspective view with time progressing to the right. We begin with an initial data qubit state that has known parities in one stabilizer basis (here, Z). We show example errors that manifest in detection pairs: a Z error (red) on a data qubit (spacelike pair), a measurement error (purple) on a measure qubit (timelike pair), an X error (blue) during the CZ gates (spacetimelike pair) and a measurement error (green) on a data qubit (detected in the final inferred Z parities). **b**, Detection probability for each stabilizer over a 25-cycle distance-5 experiment (50,000 repetitions). Darker lines: average over all stabilizers with the same weight. There are fewer detections at timestep $t = 0$ because there is no preceding syndrome extraction, and at $t = 25$ because the final parities are calculated from data qubit measurements directly. QEC, quantum error correction. **c**, Detection probability heatmap, averaging over $t = 1$ to 24. **d**,**e**, Similar to **b**,**c** for four separate distance-3 experiments covering the four quadrants of the distance-5 code. **f**,**g**, Similar to **b**,**c** using a simulation with Pauli errors plus leakage, crosstalk and stray interactions (Pauli+). **h**, Bar chart summarizing the detection correlation matrix p_{ij}, comparing the distance-5 experiment from **b** to the simulation in **f** (Pauli+) and a simpler simulation with only Pauli errors. We aggregate four groups of correlations: timelike pairs;

spacelike pairs; spacetimelike pairs expected for Pauli noise; and spacetimelike pairs unexpected for Pauli noise (Unexp.), including correlations over two timesteps. Each bar shows a mean and standard deviation of correlations from a 25-cycle, 50,000-repetition dataset.

Our stabilizer circuits contain a few modifications to the standard gate sequence described above (see Supplementary Information), including phase corrections to correct for unintended qubit frequency shifts and dynamical decoupling gates during qubit idles[43]. We also remove certain Hadamard gates to implement the *ZXXZ* variant of the surface code[44,45], which helps symmetrize the *X*- and *Z*-basis logical error rates. Finally, during initialization, the data qubits are prepared into randomly selected bitstrings. This ensures that we do not preferentially measure even parities in the first few cycles of the code, which could artificially lower logical error rates owing to bias in measurement error (see Supplementary Information).

Error detectors

After initialization, parity measurements should produce the same value in each cycle, up to known flips applied by the circuit. If we compare a parity measurement to the corresponding measurement in the preceding cycle and their values are inconsistent, a detection event has occurred, indicating an error. We refer to these comparisons as detectors.

The detection event probabilities for each detector indicate the distribution of physical errors in space and time while running the surface code. In Fig. 2, we show the detection event probabilities in the distance-5 code (Fig. 2b,c) and the distance-3 codes (Fig. 2d,e) running for 25 cycles, as measured over 50,000 experimental instances. For the weight-4 stabilizers, the average detection probability is 0.185 ± 0.018 (1σ) in the distance-5 code and 0.175 ± 0.017 averaged over the distance-3 codes. The weight-2 stabilizers interact with fewer qubits and hence detect fewer errors. Correspondingly, they yield a lower average detection probability of 0.119 ± 0.012 in the distance-5 code and 0.115 ± 0.008 averaged over the distance-3 codes. The relative consistency between code distances suggests that growing the lattice does not substantially increase the component error rates during error correction.

The average detection probabilities exhibit a relative rise of 12% for distance-5 and 8% for distance-3 over 25 cycles, with a typical characteristic risetime of roughly 5 cycles (see Supplementary Information). We attribute this rise to data qubits leaking into non-computational excited states and anticipate that the inclusion of leakage-removal techniques on data qubits would help to mitigate this rise[42,46,47,48]. We reason that the greater increase in detection probability in the distance-5 code is due to increased stray interactions or leakage from simultaneously operating more gates and measurements.

We test our understanding of the physical noise in our system by comparing the experimental data to a simulation. We begin with a depolarizing noise simulation based on the component error information in Fig. 1c, and then extend to a Pauli simulation with qubit-specific T_1 and $T_{2,\mathrm{CPMG}}$, transitions to leaked states, and stray interactions between qubits during CZ gates (see Supplementary Information). We refer to this simulation as Pauli+. Figure 2f shows that this second simulator accurately predicts the average detection probabilities, finding 0.180 ± 0.013 for the weight-4 stabilizers and 0.116 ± 0.011 for the weight-2 stabilizers, with average detection probabilities increasing 7% over 25 cycles (distance-5).

Understanding errors through correlations

We next examine pairwise correlations between detection events, which give us fine-grained information about which types of error are occurring during error correction. Figure 2a illustrates a few examples of pairwise detections that are generated by X or Z errors in the surface code. Measurement and reset errors are detected by the same stabilizer in two consecutive cycles, which we classify as a timelike pair. Data qubits may experience an X (Z) error while idling during measurement that is detected by its neighbouring Z (X) stabilizers in the same cycle, forming a spacelike pair. Errors during CZ gates may cause a variety of pairwise detections to occur, including spacetimelike pairs that are separated in both space and time. More complex clusters of detection events arise when a Y error occurs, which generates detection events for both X and Z errors.

To estimate the probability for each detection event pair from our data, we compute an appropriately normalized correlation p_{ij} between detection events occurring on any two detectors i and j (refs. [42,49]; see Supplementary Information). In Fig. 2h, we show the estimated probabilities for experimental and simulated distance-5 data, aggregated and averaged according to the different classes of pairs. In addition to the expected pairs, we also quantify how often detection pairs occur that are unexpected in a local depolarizing circuit model. Overall, the Pauli simulation systematically underpredicts these probabilities compared to experimental data, whereas the Pauli+ simulation is closer and predicts the presence of unexpected pairs, which we surmise are related to leakage and stray interactions. These errors can be especially harmful to the surface code because they can generate multiple detection events distantly separated in space or time, which a decoder might wrongly interpret as multiple independent component errors. We expect that mitigating leakage and stray interactions will become increasingly important as error rates decrease.

Decoding and logical error probabilities

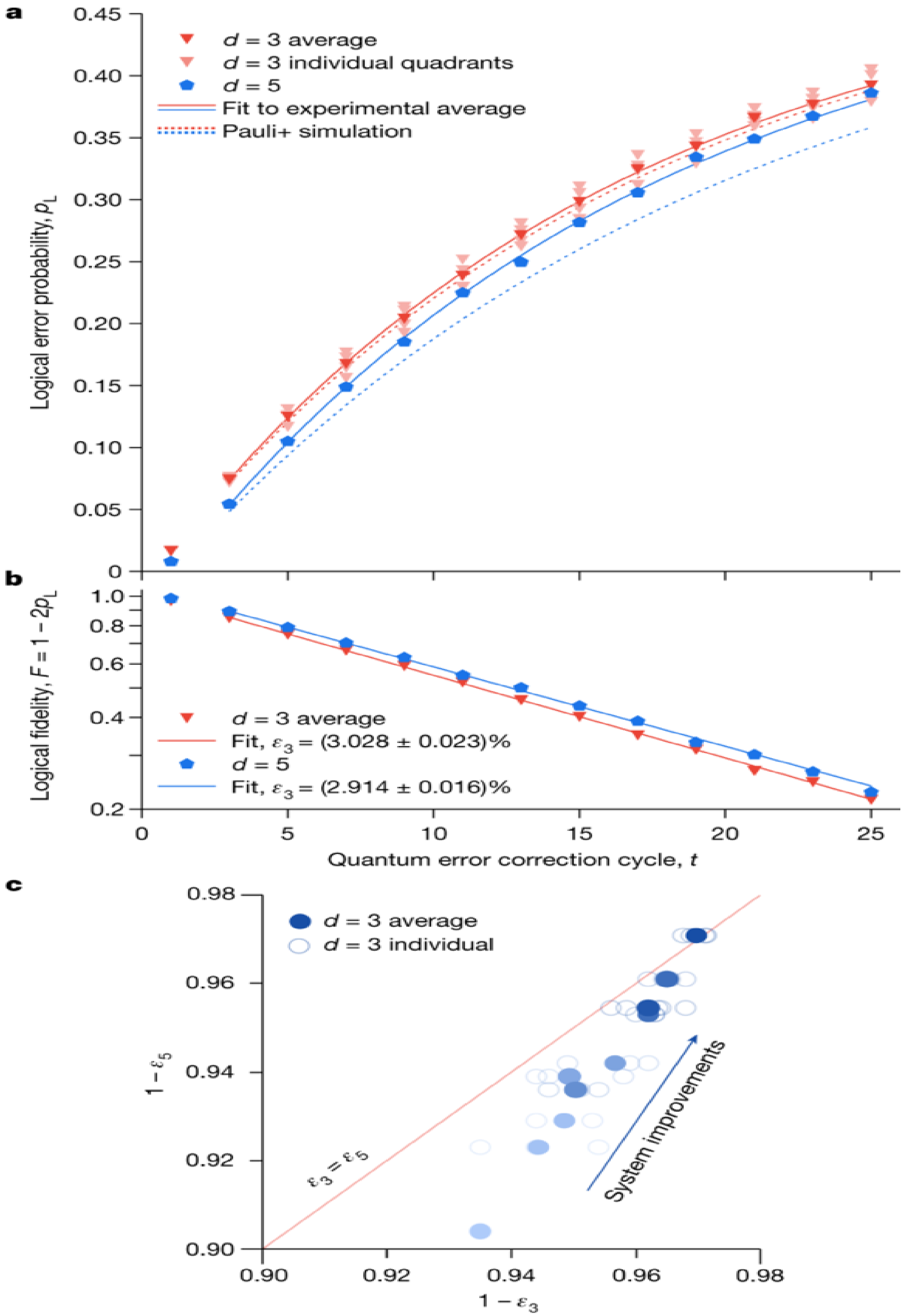
We next examine the logical performance of our surface code qubits. To infer the error-corrected logical measurement, the decoder requires a probability model for physical error events. This information may be expressed as an error hypergraph: detectors are vertices, physical error mechanisms are hyperedges connecting the detectors they trigger, and each hyperedge is assigned its corresponding error mechanism probability. We use a generalization of p_{ij} to determine these probabilities[42,50].

Given the error hypergraph, we implement two different decoders: belief-matching, an efficient combination of belief propagation and minimum-weight perfect matching[51]; and tensor network decoding, a slow but accurate approximate maximum-likelihood decoder. The belief-matching decoder first runs belief propagation on the error hypergraph to update hyperedge error probabilities based on nearby detection events[51,52]. The updated error hypergraph is then decomposed into a pair of disjoint error graphs, one each for X and Z errors[31]. These graphs are decoded efficiently using minimum-weight perfect matching[53] to select a single probable set of errors.

By contrast, a maximum-likelihood decoder considers all possible sets of errors consistent with the detection events, splits them into two groups on the basis of whether they flip the logical measurement, and chooses the group with the greater total likelihood. The two likelihoods are each expressed as a tensor network contraction[51,54,55] that exhaustively sums the probabilities of all sets of errors within each group. We can contract the network approximately, and verify that the approximation converges. This yields a decoder that is nearly optimal given the hypergraph error priors, but is considerably slower. Further improvements could come from a more accurate prior, or by incorporating more fine-grained measurement information[47,56].

Figure 3 shows a comparison of the logical error performance of the distance-3 and distance-5 codes using the approximate maximum-likelihood decoder. As the *ZXXZ* variant of the surface code symmetrizes the X and Z bases, differences between the two bases' logical error per cycle are small and attributable to spatial variations in physical error rates. Thus, for visual clarity, we report logical error probabilities averaged between the X and Z basis; the full dataset may be found in the Supplementary Information. Note that we do not post-select on leakage or high-energy events to capture the effects of realistic non-idealities on logical performance. Over all 25 cycles of error correction, the distance-5 code realizes lower logical error probabilities p_{L} than the average of the subset distance-3 codes.

Fig. 3: Logical error reduction.



a
d = 3 average
d = 3 individual quadrants
d = 5
Fit to experimental average
Pauli+ simulation
Logical error probability, pL
b
Logical fidelity, F = 1 − 2pL
d = 3 average
Fit, ε3 = (3.028 ± 0.023)%
d = 5
Fit, ε3 = (2.914 ± 0.016)%
Quantum error correction cycle, t
c
d = 3 average
d = 3 individual
ε3 = ε5
System improvements
1 − ε5
1 − ε3

a, Logical error probability p_L versus cycle comparing distance-5 (blue) to distance-3 (pink: four separate quadrants, red: average), all averaged over Z_L and X_L. Each individual data point represents 100,000 repetitions. Solid line: fit to experimental average, $t = 3$ to 25 (see main text). Dotted line: comparison to Pauli+ simulation. **b**, Logical fidelity $F = 1 - 2p_L$ versus cycle, semilog plot. The datapoints and fits are the experimental averages and fits from **a**. **c**, Summary of experimental progression comparing logical error per cycle ε_d (specifically plotting $1 - \varepsilon_d$) between distance-3 and distance-5, for which system improvements lead to faster improvement for distance-5 (see main text). Each open circle is a comparison to a specific distance-3 code, and filled circles average over several distance-3 codes measured in the same session. Markers are coloured chronologically from light to dark. Typical 1σ statistical and fit uncertainty is 0.02%, smaller than the points.

We fit the logical fidelity $F = 1 - 2p_L$ to an exponential decay. We start the fit at $t = 3$ to avoid two phenomena that advantage the larger code: the lower detection probability during the first cycle relative to subsequent cycles (Fig. 2b,d), and the higher effective threshold caused by the confinement of errors to thin time slices in few-cycle experiments[31]. We obtain a logical error per cycle $\varepsilon_5 = (2.914 \pm 0.016)\%$ (1σ statistical and fit uncertainty) for the distance-5 code, compared to an average of $\varepsilon_3 = (3.028 \pm 0.023)\%$ for the subset distance-3 codes, a relative error reduction of about 4%. When decoding with the faster belief-matching decoder, we fit a logical error per cycle of $(3.056 \pm 0.015)\%$ for the distance-5 code, compared to an average of $(3.118 \pm 0.025)\%$ for the distance-3 codes, a relative error reduction of about 2%. We note that the distance-5 logical error per cycle is slightly higher than those of two of the distance-3 codes individually, and that leakage accumulation may cause distance-5 performance to degrade faster than that of distance-3 as logical error probability approaches 50%.

In principle, the logical performance of a distance-5 code should improve faster than that of a distance-3 code as physical error rates decrease[33]. Over time, we improved our physical error rates, for example by optimizing single- and two-qubit gates, measurement and data qubit idling (see Supplementary Information). In Fig. 3c, we show the corresponding

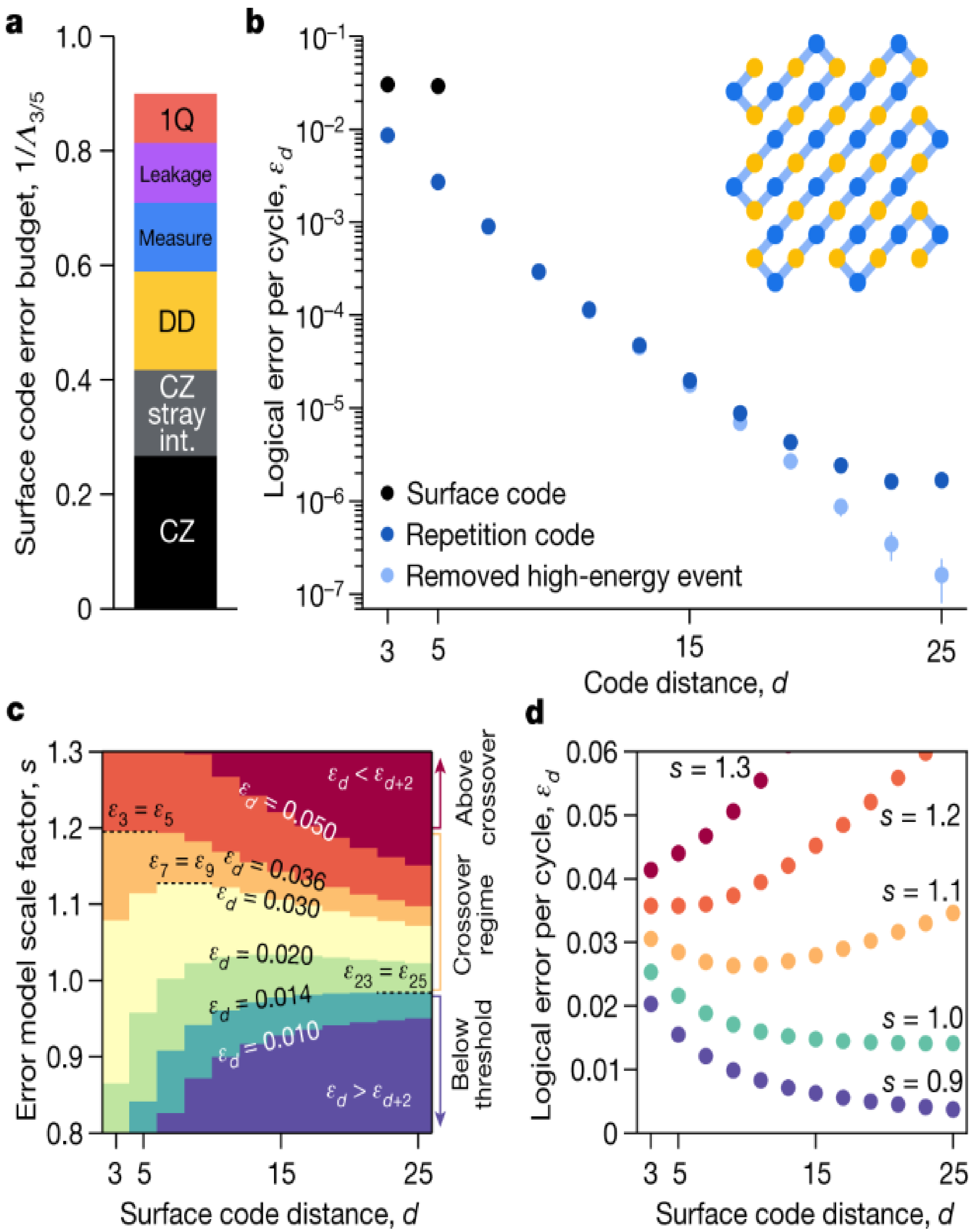
performance progression of distance-5 and distance-3 codes. The larger code improved about twice as fast until finally overtaking the smaller code, validating the benefit of increased-distance protection in practice.

To understand the contributions of individual components to our logical error performance, we follow ref. [42] and simulate the distance-5 and distance-3 codes while varying the physical error rates of the various circuit components. As the logical-error-suppression factor

$${{\Lambda }}_{d/(d+2)}={\varepsilon }_{d}/{\varepsilon }_{d+2} \quad (1)$$

is approximately inversely proportional to the physical error rate, we can budget how much each physical error mechanism contributes to $1/\Lambda_{3/5}$ (as shown in Fig. 4a) to assess scaling. This error budget shows that CZ error and data qubit decoherence during measurement and reset are dominant contributors.

Fig. 4: Towards algorithmically relevant error rates.



a, Estimated error budget for the surface code, based on component errors (see Fig. 1c) and Pauli+ simulations. $\Lambda_{3/5} = \varepsilon_3/\varepsilon_5$. CZ, contributions from CZ error (excluding leakage and stray interactions). CZ stray int., CZ error from unwanted interactions. DD, dynamical decoupling (data qubit idle error

during measurement and reset). Measure, measurement and reset error. Leakage, leakage during CZs and due to heating. 1Q, single-qubit gate error. **b**, Logical error for repetition codes. Inset: schematic of the distance-25 repetition code, using the same data and measure qubits as the distance-5 surface code. Smaller codes are subsampled from the same distance-25 data[42]. A high-energy event resulted in an apparent error floor around 10^{-6}. After removing the instances nearby (light blue), error decreases more rapidly with code distance. The dataset has 50 cycles, 5×10^5 repetitions. We also plot the surface code error per cycle from Fig. 3b in black. **c**, Contour plot of simulated surface code logical error per cycle ε_d as a function of code distance d and a scale factor s on the error model in Fig. 1c (Pauli simulation, $s = 1.0$ corresponds to the current device error model). **d**, Horizontal slices from **c**, each for a value of error-model scale factor s. $s = 1.3$ is above threshold (larger codes are worse), and $s = 1.2$ to 1.0 represent the crossover regime, for which progressively larger codes get better until a turnaround. $s = 0.9$ is below threshold (larger codes are better).

Algorithmically relevant error rates

Even as known error sources are suppressed in future devices, new dominant error mechanisms may arise as lower logical error rates are realized. To test the behaviour of codes with substantially lower error rates, we use the bit-flip repetition code, a one-dimensional version of the surface code. The bit-flip repetition code does not correct for phase-flip errors and is thus unsuitable for quantum algorithms. However, correcting only bit-flip errors allows it to achieve much lower logical error probabilities.

Without post-selection, we achieve a logical error per cycle of $(1.7 \pm 0.3) \times 10^{-6}$ using a distance-25 repetition code decoded with minimum-weight perfect matching (Fig. 4b). We attribute many of these logical errors in the higher-distance codes to a high-energy impact, which can temporarily impart widespread correlated errors to the system[57]. These events may be identified by spikes in detection event counts[42], and such error mechanisms must be mitigated for scalable quantum error correction to succeed. In this case, there was one such event; after removing it (0.15% of trials), we observe a logical error per cycle of $(1.6 \pm 0.8) \times 10^{-7}$ (see Supplementary Information).

The repetition code results demonstrate that low logical error rates are possible in a superconducting system, but finding and mitigating highly correlated errors such as cosmic ray impacts will be an important area of research moving forwards.

Towards large-scale quantum error correction

To understand how our surface code results project forwards to future devices, we simulate the logical error performance of surface codes ranging from distance-3 to 25, while also scaling the physical error rates shown in Fig. 1c. For efficiency, the simulation considers only Pauli errors. Figure 4c,d illustrates the contours of this parameter space, which has three distinct regions. When the physical error rate is high (for example, the initial runs of our surface code in Fig. 3c), logical error probability increases with increasing system size ($\varepsilon_{d+2} > \varepsilon_d$). On the other hand, low physical error rates show the desired exponential suppression of logical error ($\varepsilon_{d+2} < \varepsilon_d$). This threshold behaviour can be subtle[58], and there exists a crossover regime in which, owing to finite-size effects, increasing system size initially suppresses the logical error per cycle before later increasing it. We believe our experiment lies in this regime.

Although our device is close to threshold, reaching algorithmically relevant logical error rates with manageable resources will require an error-suppression factor $\Lambda_{d/(d+2)} \gg 1$. On the basis of the error budget and simulations in Fig. 4, we estimate that component performance must improve by at least 20% to move below threshold, and substantially improve beyond that to achieve practical scaling. However, these projections rely on simplified models and must be validated experimentally, testing larger code sizes with longer durations to eventually realize the desired logical performance. This work demonstrates the first step in that process, suppressing logical errors by scaling a quantum error-correcting code—the foundation of a fault-tolerant quantum computer.

Data availability

The data that support the findings of this study are available at https://doi.org/10.5281/zenodo.6804040.

References

1. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493 (1995).

2. Gottesman, D. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, California Institute of Technology (1997).

3. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).

4. Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Rev.* **41**, 303–332 (1999).

5. Farhi, E. et al. A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science* **292**, 472–475 (2001).

6. Biamonte, J. et al. Quantum machine learning. *Nature* **549**, 195–202 (2017).

7. Lloyd, S. Universal quantum simulators. *Science* **273**, 1073–1078 (1996).

8. Aspuru-Guzik, A., Dutoi, A. D., Love, P. J. & Head-Gordon, M. Simulated quantum computation of molecular energies. *Science* **309**, 1704–1707 (2005).

9. Reiher, M., Wiebe, N., Svore, K. M., Wecker, D. & Troyer, M. Elucidating reaction mechanisms on quantum computers. *Proc. Natl Acad. Sci. USA* **114**, 7555–7560 (2017).

10. Gidney, C. & Ekera, M. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum* **5**, 433 (2021).

11. Kivlichan, I. D. et al. Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via trotterization. *Quantum* **4**, 296 (2020).

12. Ballance, C., Harty, T., Linke, N., Sepiol, M. & Lucas, D. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Phys. Rev. Lett.* **117**, 060504 (2016).

13. Huang, W. et al. Fidelity benchmarks for two-qubit gates in silicon. *Nature* **569**, 532–536 (2019).

14. Rol, M. et al. Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits. *Phys. Rev. Lett.* **123**, 120502 (2019).

15. Jurcevic, P. et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.* **6**, 025020 (2021).

16. Foxen, B. et al. Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms. *Phys. Rev. Lett.* **125**, 120504 (2020).

17. Wu, Y. et al. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* **127**, 180501 (2021).

18. Knill, E., Laflamme, R. & Zurek, W. H. Resilient quantum computation. *Science* **279**, 342–345 (1998).

19. Aharonov, D. & Ben-Or, M. Fault-tolerant quantum computation with constant error rate. *SIAM J. Comput.* **38**, 1207–1282 (2008).

20. Egan, L. et al. Fault-tolerant control of an error-corrected qubit. *Nature* **598**, 281–286 (2021).

21. Ryan-Anderson, C. et al. Realization of real-time fault-tolerant quantum error correction. *Phys. Rev. X* **11**, 041058 (2021).

22. Abobeih, M. et al. Fault-tolerant operation of a logical qubit in a diamond quantum processor. *Nature* **606**, 884–889 (2022).

23. Sundaresan, N. et al. Matching and maximum likelihood decoding of a multi-round subsystem quantum error correction experiment. Preprint at https://arXiv.org/abs/2203.07205 (2022).

24. Krinner, S. et al. Realizing repeated quantum error correction in a distance-three surface code. *Nature* **605**, 669–674 (2022).

25. Zhao, Y. et al. Realization of an error-correcting surface code with superconducting qubits. *Phys. Rev. Lett.* **129**, 030501 (2022).

26. Ofek, N. et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441–445 (2016).

27. Flühmann, C. et al. Encoding a qubit in a trapped-ion mechanical oscillator. *Nature* **566**, 513–517 (2019).

28. Campagne-Ibarcq, P. et al. Quantum error correction of a qubit encoded in grid states of an oscillator. *Nature* **584**, 368–372 (2020).

29. Grimm, A. et al. Stabilization and operation of a Kerr-cat qubit. *Nature* **584**, 205–209 (2020).

30. Kitaev, A. Y. Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303**, 2–30 (2003).

31. Dennis, E., Kitaev, A., Landahl, A. & Preskill, J. Topological quantum memory. *J. Math. Phys.* **43**, 4452–4505 (2002).

32. Raussendorf, R. & Harrington, J. Fault-tolerant quantum computation with high threshold in two dimensions. *Phys. Rev. Lett.* **98**, 190504 (2007).

33. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).

34. Satzinger, K. et al. Realizing topologically ordered states on a quantum processor. *Science* **374**, 1237–1241 (2021).

35. Horsman, C., Fowler, A. G., Devitt, S. & Meter, R. V. Surface code quantum computing by lattice surgery. *New J. Phys.* **14**, 123011 (2012).

36. Fowler, A. G. & Gidney, C. Low overhead quantum computation using lattice surgery. Preprint at https://arXiv.org/abs/1808.06709 (2018).

37. Litinski, D. A game of surface codes: large-scale quantum computing with lattice surgery. *Quantum* **3**, 128 (2019).

38. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).

39. Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).

40. Neill, C. *A Path towards Quantum Supremacy with Superconducting Qubits*. PhD thesis, Univ. California Santa Barbara (2017).

41. Yan, F. et al. Tunable coupling scheme for implementing high-fidelity two-qubit gates. *Phys. Rev. Appl.* **10**, 054062 (2018).

42. Chen, Z. et al. Exponential suppression of bit or phase errors with cyclic error correction. *Nature* **595**, 383–387 (2021).

43. Kelly, J. et al. Scalable in situ qubit calibration during repetitive error detection. *Phys. Rev. A* **94**, 032321 (2016).

44. Wen, X.-G. Quantum orders in an exact soluble model. *Phys. Rev. Lett.* **90**, 016803 (2003).

45. Bonilla Ataides, J. P., Tuckett, D. K., Bartlett, S. D., Flammia, S. T. & Brown, B. J. The XZZX surface code. *Nat. Commun.* **12**, 2172 (2021).

46. Aliferis, P. & Terhal, B. M. Fault-tolerant quantum computation for local leakage faults. *Quantum Inf. Comput.* **7**, 139–156 (2007).

47. Suchara, M., Cross, A. W. & Gambetta, J. M. Leakage suppression in the toric code. *Proc. 2015 IEEE International Symposium on Information Theory (ISIT)* 1119–1123 (2015).

48. McEwen, M. et al. Removing leakage-induced correlated errors in superconducting quantum error correction. *Nat. Commun.* **12**, 1761 (2021).

49. Spitz, S. T., Tarasinski, B., Beenakker, C. W. & O'Brien, T. E. Adaptive weight estimator for quantum error correction in a time-dependent environment. *Adv. Quantum Technol.* **1**, 1800012 (2018).

50. Chen, E. H. et al. Calibrated decoders for experimental quantum error correction. *Phys. Rev. Lett.* **128**, 110504 (2022).

51. Higgott, O., Bohdanowicz, T. C., Kubica, A., Flammia, S. T. & Campbell, E. T. Fragile boundaries of tailored surface codes and improved decoding of circuit-level noise. Preprint at https://arXiv.org/abs/2203.04948 (2022).

52. Criger, B. & Ashraf, I. Multi-path summation for decoding 2D topological codes. *Quantum* **2**, 102 (2018).

53. Fowler, A. G., Whiteside, A. C. & Hollenberg, L. C. Towards practical classical processing for the surface code. *Phys. Rev. Lett.* **108**, 180501 (2012).

54. Bravyi, S., Suchara, M. & Vargo, A. Efficient algorithms for maximum likelihood decoding in the surface code. *Phys. Rev. A* **90**, 032326 (2014).

55. Chubb, C. T. & Flammia, S. T. Statistical mechanical models for quantum codes with correlated noise. *Ann. Inst. Henri Poincaré D* **8**, 269–321 (2021).

56. Pattison, C. A., Beverland, M. E., da Silva, M. P. & Delfosse, N. Improved quantum error correction using soft information. Preprint at https://arXiv.org/abs/2107.13589 (2021).

57. McEwen, M. et al. Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits. *Nat. Phys.* **18**, 107–111 (2022).

58. Stephens, A. M. Fault-tolerant thresholds for quantum error correction with the surface code. *Phys. Rev. A* **89**, 022321 (2014).

59. Emerson, J., Alicki, R. & Życzkowski, K. Scalable noise estimation with random unitary operators. *J. Opt. B* **7**, S347 (2005).

Acknowledgements

We are grateful to S. Brin, S. Pichai, R. Porat, J. Dean, E. Collins and J. Yagnik for their executive sponsorship of the Google Quantum AI team, and for their continued engagement and support. A portion of this work was performed in the University of California, Santa Barbara Nanofabrication Facility, an open access laboratory. J.M. acknowledges support from the National Aeronautics and Space Administration (NASA) Ames Research Center (NASA-Google SAA 403512), NASA Advanced Supercomputing Division for access to NASA high-performance computing systems, and NASA Academic Mission Services (NNA16BD14C). D.B. is a CIFAR Associate Fellow in the Quantum Information Science Program.

Author information

Authors and Affiliations

1. Google Research, Mountain View, CA, USA

   Rajeev Acharya, Igor Aleiner, Richard Allen, Trond I. Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Juan Atalaya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Joao Basso, Andreas Bengtsson, Sergio Boixo, Gina Bortoli, Alexandre Bourassa, Jenna Bovaird, Leon Brill, Michael Broughton, Bob B. Buckley, David A. Buell, Tim Burger, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Josh Cogan, Roberto Collins, Paul Conner, William Courtney, Alexander L.

Crook, Ben Curtin, Dripto M. Debroy, Alexander Del Toro Barba, Sean Demura, Andrew Dunsworth, Daniel Eppens, Catherine Erickson, Lara Faoro, Edward Farhi, Reza Fatemi, Leslie Flores Burgos, Ebrahim Forati, Austin G. Fowler, Brooks Foxen, William Giang, Craig Gidney, Dar Gilboa, Marissa Giustina, Alejandro Grajales Dau, Jonathan A. Gross, Steve Habegger, Michael C. Hamilton, Matthew P. Harrigan, Sean D. Harrington, Oscar Higgott, Jeremy Hilton, Markus Hoffmann, Sabrina Hong, Trent Huang, Ashley Huff, William J. Huggins, Lev B. Ioffe, Sergei V. Isakov, Justin Iveland, Evan Jeffrey, Zhang Jiang, Cody Jones, Pavol Juhas, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Tanuj Khattar, Mostafa Khezri, Mária Kieferová, Seon Kim, Alexei Kitaev, Paul V. Klimov, Andrey R. Klots, Alexander N. Korotkov, Fedor Kostritsa, John Mark Kreikebaum, David Landhuis, Pavel Laptev, Kim-Ming Lau, Lily Laws, Joonho Lee, Kenny Lee, Brian J. Lester, Alexander Lill, Wayne Liu, Aditya Locharla, Erik Lucero, Fionn D. Malone, Orion Martin, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Bernardo Meurer Costa, Xiao Mi, Kevin C. Miao, Masoud Mohseni, Shirin Montazeri, Alexis Morvan, Emily Mount, Wojciech Mruczkiewicz, Ofer Naaman, Matthew Neeley, Charles Neill, Ani Nersisyan, Hartmut Neven, Michael Newman, Jiun How Ng, Anthony Nguyen, Murray Nguyen, Murphy Yuezhen Niu, Thomas E. O'Brien, Alex Opremcak, John Platt, Andre Petukhov, Rebecca Potter, Leonid P. Pryadko, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Negar Saei, Daniel Sank, Kannan Sankaragomathi, Kevin J. Satzinger, Henry F. Schurkus, Christopher Schuster, Michael J. Shearn, Aaron Shorter, Vladimir Shvarts, Jindra Skruzny, Vadim Smelyanskiy, W. Clarke Smith, George Sterling, Doug Strain, Marco Szalay, Alfredo Torres, Guifre Vidal, Benjamin Villalonga, Catherine Vollgraff Heidweiller, Theodore White, Cheng Xing, Z. Jamie Yao, Ping Yeh, Juhwan Yoo, Grayson Young, Adam Zalcman, Yaxing Zhang & Ningfeng Zhu

2. Department of Physics, Columbia University, New York, NY, USA

Igor Aleiner

3. Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA

Joseph C. Bardin

4. Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA

Michael C. Hamilton

5. Centre for Quantum Computation and Communication Technology, Centre for Quantum Software and Information, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, New South Wales, Australia

Mária Kieferová

6. Department of Physics, Institute for Quantum Information and Matter, and Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA, USA

Alexei Kitaev

7. Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA

Alexander N. Korotkov

8. USRA Research Institute for Advanced Computer Science, Mountain View, CA, USA

Jeffrey Marshall

9. QuAIL, NASA Ames Research Center, Mountain View, CA, USA

Jeffrey Marshall

10. Department of Physics, University of California, Santa Barbara, CA, USA

Matt McEwen

11. Department of Physics and Astronomy, University of California, Riverside, CA, USA

Leonid P. Pryadko

Consortia

Google Quantum AI

- Rajeev Acharya
- , Igor Aleiner
- , Richard Allen
- , Trond I. Andersen
- , Markus Ansmann
- , Frank Arute
- , Kunal Arya
- , Abraham Asfaw
- , Juan Atalaya
- , Ryan Babbush
- , Dave Bacon
- , Joseph C. Bardin
- , Joao Basso
- , Andreas Bengtsson
- , Sergio Boixo
- , Gina Bortoli
- , Alexandre Bourassa
- , Jenna Bovaird
- , Leon Brill
- , Michael Broughton
- , Bob B. Buckley
- , David A. Buell
- , Tim Burger
- , Brian Burkett
- , Nicholas Bushnell
- , Yu Chen

- , Zijun Chen
- , Ben Chiaro
- , Josh Cogan
- , Roberto Collins
- , Paul Conner
- , William Courtney
- , Alexander L. Crook
- , Ben Curtin
- , Dripto M. Debroy
- , Alexander Del Toro Barba
- , Sean Demura
- , Andrew Dunsworth
- , Daniel Eppens
- , Catherine Erickson
- , Lara Faoro
- , Edward Farhi
- , Reza Fatemi
- , Leslie Flores Burgos
- , Ebrahim Forati
- , Austin G. Fowler
- , Brooks Foxen
- , William Giang
- , Craig Gidney
- , Dar Gilboa
- , Marissa Giustina
- , Alejandro Grajales Dau
- , Jonathan A. Gross
- , Steve Habegger
- , Michael C. Hamilton
- , Matthew P. Harrigan
- , Sean D. Harrington
- , Oscar Higgott
- , Jeremy Hilton
- , Markus Hoffmann
- , Sabrina Hong
- , Trent Huang
- , Ashley Huff

- , William J. Huggins
- , Lev B. Ioffe
- , Sergei V. Isakov
- , Justin Iveland
- , Evan Jeffrey
- , Zhang Jiang
- , Cody Jones
- , Pavol Juhas
- , Dvir Kafri
- , Kostyantyn Kechedzhi
- , Julian Kelly
- , Tanuj Khattar
- , Mostafa Khezri
- , Mária Kieferová
- , Seon Kim
- , Alexei Kitaev
- , Paul V. Klimov
- , Andrey R. Klots
- , Alexander N. Korotkov
- , Fedor Kostritsa
- , John Mark Kreikebaum
- , David Landhuis
- , Pavel Laptev
- , Kim-Ming Lau
- , Lily Laws
- , Joonho Lee
- , Kenny Lee
- , Brian J. Lester
- , Alexander Lill
- , Wayne Liu
- , Aditya Locharla
- , Erik Lucero
- , Fionn D. Malone
- , Jeffrey Marshall
- , Orion Martin
- , Jarrod R. McClean
- , Trevor McCourt

- , Matt McEwen
- , Anthony Megrant
- , Bernardo Meurer Costa
- , Xiao Mi
- , Kevin C. Miao
- , Masoud Mohseni
- , Shirin Montazeri
- , Alexis Morvan
- , Emily Mount
- , Wojciech Mruczkiewicz
- , Ofer Naaman
- , Matthew Neeley
- , Charles Neill
- , Ani Nersisyan
- , Hartmut Neven
- , Michael Newman
- , Jiun How Ng
- , Anthony Nguyen
- , Murray Nguyen
- , Murphy Yuezhen Niu
- , Thomas E. O’Brien
- , Alex Opremcak
- , John Platt
- , Andre Petukhov
- , Rebecca Potter
- , Leonid P. Pryadko
- , Chris Quintana
- , Pedram Roushan
- , Nicholas C. Rubin
- , Negar Saei
- , Daniel Sank
- , Kannan Sankaragomathi
- , Kevin J. Satzinger
- , Henry F. Schurkus
- , Christopher Schuster
- , Michael J. Shearn
- , Aaron Shorter

- , Vladimir Shvarts
- , Jindra Skruzny
- , Vadim Smelyanskiy
- , W. Clarke Smith
- , George Sterling
- , Doug Strain
- , Marco Szalay
- , Alfredo Torres
- , Guifre Vidal
- , Benjamin Villalonga
- , Catherine Vollgraff Heidweiller
- , Theodore White
- , Cheng Xing
- , Z. Jamie Yao
- , Ping Yeh
- , Juhwan Yoo
- , Grayson Young
- , Adam Zalcman
- , Yaxing Zhang
- & Ningfeng Zhu

Contributions

The Google Quantum AI team conceived and designed the experiment. The theory and experimental teams at Google Quantum AI developed the data analysis, modelling and metrological tools that enabled the experiment, built the system, performed the calibrations and collected the data. The modelling was carried out jointly with collaborators outside Google Quantum AI. All authors wrote and revised the manuscript and the Supplementary Information.

Corresponding author

Correspondence to Hartmut Neven.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Barbara Terhal, Boris Varbanov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Supplementary Information

This file contains Supplementary Sections 13–21, Figs. 1–34, a full list of members of Google Quantum AI and References.

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **614**, 676–681 (2023). https://doi.org/10.1038/s41586-022-05434-1

- Received: 13 July 2022
- Accepted: 10 October 2022
- Published: 22 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05434-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

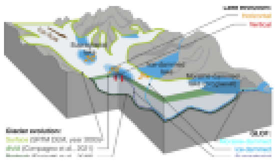
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Towards quantum computers that are robust to errors

Nature Research Briefing 22 Feb 2023

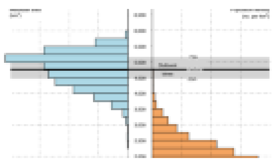
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05434-1



Future growth and decline of high mountain Asia's ice-dammed lakes and associated risk

26 August 2022

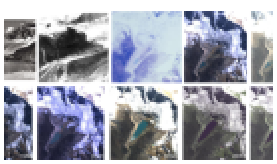
Loris Compagno, Matthias Huss, … Daniel Farinotti



Unchanged frequency of moraine-dammed glacial lake outburst floods in the Himalaya

18 March 2019

Georg Veh, Oliver Korup, … Ariane Walz



Increased outburst flood hazard from Lake Palcacocha due to human-induced glacier retreat

04 February 2021

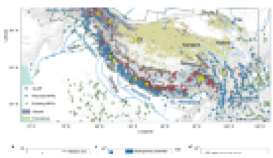
R. F. Stuart-Smith, G. H. Roe, … M. R. Allen



[Trends in the Timing and Magnitude of Ice-Jam Floods in Canada]

11 April 2018

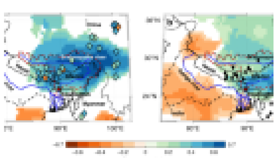
Prabin Rokaya, Sujata Budhathoki & Karl-Erich Lindenschmidt



[High Mountain Asia hydropower systems threatened by climate-driven landscape instability]

23 June 2022

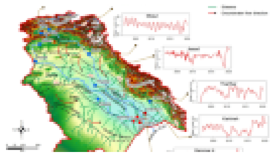
Dongfeng Li, Xixi Lu, … Tobias Bolch



[Seven centuries of reconstructed Brahmaputra River discharge demonstrate underestimated high discharge and flood hazard frequency]

26 November 2020

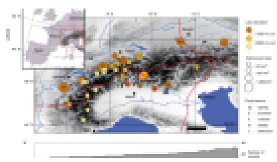
Mukund P. Rao, Edward R. Cook, … Peter J. Webster



[Buffering the impacts of extreme climate variability in the highly engineered Tigris Euphrates river system]

09 March 2022

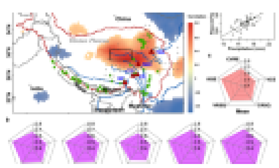
Karem Abdelmohsen, Mohamed Sultan, … Khaled H. Zahran



[Impact of warmer climate periods on flood hazard in the European Alps]

27 January 2022

B. Wilhelm, W. Rapuc, … S. B. Wirth



[Reconstructed eight-century streamflow in the Tibetan Plateau reveals contrasting regional variability and strong nonstationarity]

27 October 2022

Yenan Wu, Di Long, … Chunhong Hu

- Article
- Open Access

- Published: 15 February 2023

Less extreme and earlier outbursts of ice-dammed lakes since 1900

- Georg Veh ,
- Natalie Lützow,
- Jenny Tamm,
- Lisa V. Luna,
- Romain Hugonnet,
- Kristin Vogel,
- Marten Geertsema,
- John J. Clague &
- …
- Oliver Korup

Show authors

Nature **volume 614**, pages 701–707 (2023)

- 2920 Accesses

- 114 Altmetric

- Metrics details

Abstract

Episodic failures of ice-dammed lakes have produced some of the largest floods in history, with disastrous consequences for communities in high mountains[1,2,3,4,5,6,7]. Yet, estimating changes in the activity of ice-dam failures through time remains controversial because of inconsistent regional flood databases. Here, by collating 1,569 ice-dam failures in six major mountain regions, we systematically assess trends in peak discharge, volume, annual timing and source elevation between 1900 and 2021. We

show that extreme peak flows and volumes (10 per cent highest) have declined by about an order of magnitude over this period in five of the six regions, whereas median flood discharges have fallen less or have remained unchanged. Ice-dam floods worldwide today originate at higher elevations and happen about six weeks earlier in the year than in 1900. Individual ice-dammed lakes with repeated outbursts show similar negative trends in magnitude and earlier occurrence, although with only moderate correlation to glacier thinning[8]. We anticipate that ice dams will continue to fail in the near future, even as glaciers thin and recede. Yet widespread deglaciation, projected for nearly all regions by the end of the twenty-first century[9], may bring most outburst activity to a halt.

Main

One of the major hazards caused by ongoing glacier retreat are sudden floods resulting from the failure of unstable glacier dams[4,5,6,7]. These ice dams temporarily and periodically impound rainfall and meltwater that can be released abruptly following dam flotation, subglacial tunnel enlargement or mechanical collapse[10]. Such glacier lake outburst floods (GLOFs) initiate hazard cascades with numerous impacts downstream. Historic ice-dammed GLOFs have travelled hundreds of kilometres, entraining sediment along their runout paths; widening river channels by bank erosion; triggering landslides from undercut and destabilized hillslopes; and covering valley floors with metre-thick sheets of sediment, ice and woody debris[11,12,13,14]. Extreme GLOF discharges can raise local river levels by several metres, increasing bed shear stresses and stream power values sufficiently to incise bedrock[15]. Floods from the largest lake outbursts can last for several days and destroy local flood control systems[7,16]. In many mountain regions, settlements, tourism, forestry and mining advanced into high mountain valleys during the twentieth century[17] and thus closer to the source areas of GLOFs. As a consequence, ice-dam failures have caused many hundreds of fatalities, damaged infrastructure and farmland worth hundreds of millions of dollars, and disrupted transportation and communication routes[1].

Despite the hazard posed by ice-dammed lakes, temporal trends in the magnitude and timing of their failures remain poorly understood in most

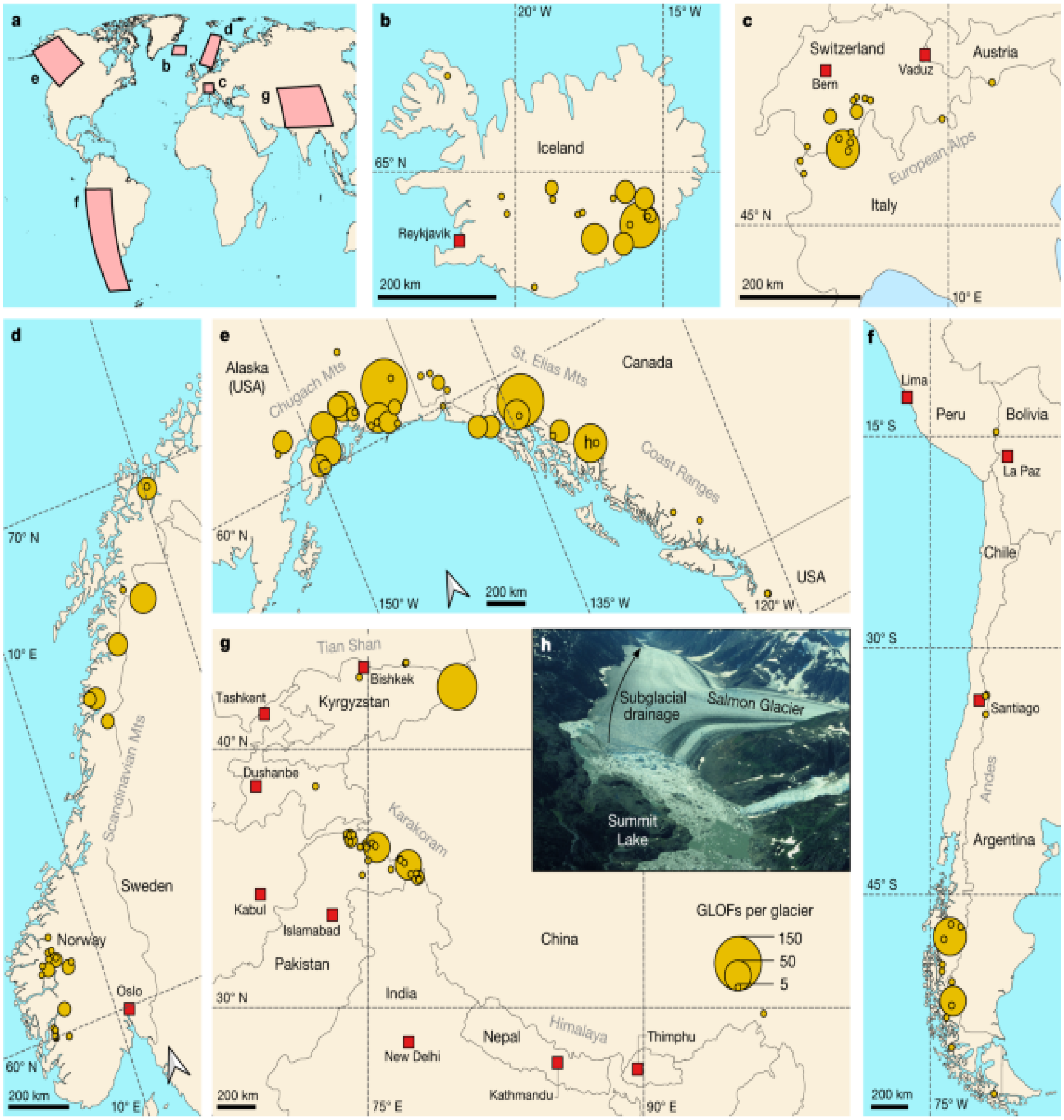
mountain regions. We attribute the limited knowledge of contemporary GLOF activity to the largely unsystematic and regionally inconsistent record of GLOFs. How strongly, or whether at all, regional GLOF activity responds to accelerating glacier melt therefore remains an open question. Melting glacier margins create new space for ice-dammed lakes to form and grow[13,14,18]. Thus, in the future, ice dams might store an increasing amount of meltwater, causing larger floods if they fail. However, the few detailed case studies of ice-dammed lakes, particularly in Iceland and northwest North America[6,11,14,19], suggest the opposite. There, individual ice-dammed lakes fill and empty repeatedly, releasing gradually smaller flood volumes over time. The edges of glacier dams might begin to float on the adjacent lake once a critical lake level is reached, suggesting that a gradually thinning glacier dam requires a lower threshold to initiate subglacial drainage[6,20,21]. Within this cycle, the timing of GLOFs, that is, the day of the year when a lake begins to drain, is likely to shift, given that thinner dams may require a shorter period of time to refill to their maximum storage capacity[14,22]. This cycle of GLOFs from a given lake may end when the ice dam has decayed to the point where it can no longer impound water[20,21]. Yet, new cycles of GLOFs might originate from higher elevations with ongoing exposure of ice-free areas at glacier margins[23].

A database of historic ice-dam failures

Here we study the core components of the GLOF cycle by examining historical changes in GLOF magnitude, timing and source elevation, and their relationships to glacier thinning. We focus on the six most glaciated mountain regions on Earth, that is, northwest North America, High Mountain Asia, the Andes, Iceland, Scandinavia and the European Alps (Fig. 1). In reviewing 446 sources of information, we found a total of 1,569 dated outbursts from 186 ice-dammed lakes in the period 1900–2021 (Methods). In 64% of all cases, we had access to primary sources such as original research papers, official documentation from local agencies, reports in newspapers and social media platforms, and written correspondence with local eyewitnesses and experts. Secondary sources (36%) include review articles or data in previously published databases, for which we could not access the primary source directly. For each GLOF, we recorded the name of

the source lake and its parent glacier; the year and, if available, the month and day of occurrence; the elevation of the lake surface; and any previously gauged or estimated values of outburst peak discharge Q_p and flood volume V_0. Using Bayesian hierarchical quantile regression models, we estimated global and regional trends in reported median and extreme discharges, defined as GLOFs exceeding the 50th and 90th percentiles of the conditional distributions of reported Q_p and V_0. We also used regression models to assess changes in the regional timing and the median source elevation of ice-dam failures (Methods). We assessed the sensitivity of the posterior trends in GLOF magnitudes learned from data since 1900 against the period since 1990, as regional GLOF reporting improved in the late twentieth century[24] (Extended Data Fig. 1).

Fig. 1: Map of dated GLOF source locations in the six study regions.



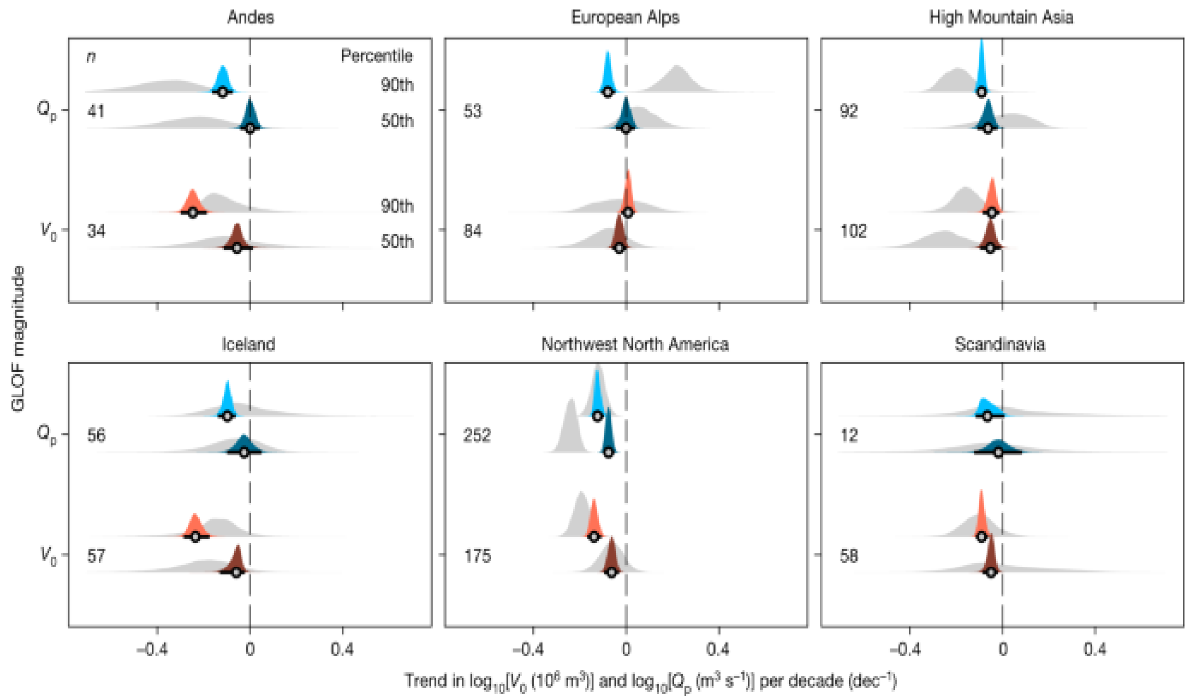
a, Global overview. **b**, Iceland. **c**, European Alps. **d**, Scandinavia. **e**, Northwest North America. **f**, Andes. **g**, High Mountain Asia, including legend. We created all maps in **a**–**g** using the free software QGIS V3.22.6 (https://www.qgis.org/). Generalized country boundaries and locations of cities are available through the online ArcGIS hub hosted by Esri Data and Maps (https://hub.arcgis.com/). **h**, Typical setting of an ice-dammed lake: Summit Lake dammed by Salmon Glacier, British Columbia, Canada (see location in **e**). Oblique view shows the situation on 11 August 1989, when the lake was approximately 1-km wide. Summit Lake has been draining

beneath Salmon Glacier (arrow) since 1961, and only a small lake exists today (Fig. 5c). Photograph by John J. Clague. Mts, Mountains.

Less extreme GLOFs since 1900

Our inventory shows that the median reported discharge and volume released by ice-dammed lakes between 1900 and 2021 were $76{7}_{-756}^{+9,761}\,{{\rm{m}}}^{3}\,{{\rm{s}}}^{-1}$ and $5{1}_{-50}^{+1,100}\times 1{0}^{6}\,{{\rm{m}}}^{3}$, respectively (median and 95% highest density interval (HDI)). Runoff peaks more than ten times higher than this global median occurred in High Mountain Asia where glacier dams collapsed suddenly after phases of advance[3,7]. Dam failures following fjord closures in the Andes[25] and Alaska[2] produced floods with volumes exceeding 10^9 m^3. However, our quantile regression models suggest that extreme GLOFs have weakened over the past 120 yr (Fig. 2 and Extended Data Figs. 2–4). The 90th percentile decreased by at least an order of magnitude in five regions for Q_p, and in four regions for V_0. The Andes and northwest North America had a decline of almost two orders of magnitude in the extremes of Q_p, whereas the highest V_0 dropped even more in the Andes and Iceland. Only Q_p in Scandinavia and V_0 in the European Alps had unchanged trends in the extremes. The regional medians remained largely stable since 1900: Q_p decreased credibly only in High Mountain Asia and northwest North America, without clear trends elsewhere. Four regions had a credible decrease in median V_0, yet in most cases trends in median V_0 were smaller than those of the extremes (Fig. 2 and Extended Data Fig. 3). A trend towards smaller floods is evident also for individual lakes with repeat (at least five) outbursts: hierarchical regression models show credible decreases in median Q_p in 23% (5 of 22), and in median V_0 in 31% (8 of 26) of all cases. Only one lake had a credible increase in Q_p or V_0 over the GLOF cycle (Extended Data Fig. 5).

Fig. 2: Trends of ice-dammed GLOF discharges between 1900 and 2021.



Posterior trends from Bayesian quantile regression of reported V_0 (orange) and Q_p (blue) with time. Dark and light colours are trends for the median and 90th percentile (that is, 10% largest), respectively, for 1900–2021; grey densities are trends for 1990–2021. Black lines are 95% HDIs and circles are medians of the posterior distributions. Trends refer to $\log_{10}$-transformed values of V_0 and Q_p, such that the posterior densities show the decadal trend in orders of magnitudes. Numbers (*n*) on the left are sample sizes for 1900–2021.

All regions had more abundant reported GLOFs since 1990 compared with earlier decades (Extended Data Fig. 1). These changes in data density probably reflect a more consistent documentation of GLOFs[24], such that an unknown number of lake outbursts may have gone unnoticed early in the study period. Reducing the observation period to the past three decades (1990–2021), a period with greater research activity, yields broader posterior trend estimates as fewer data points enter the model (Fig. 2). However, in almost all regions, the trends obtained from shorter periods are consistent with the long-term trends, even though based on less than two-thirds of the total data (Extended Data Figs. 2 and 3). Trends of extreme Q_p remain credibly negative in the Andes, High Mountain Asia and northwest North America, whereas median trends are unchanged in five of the six regions.

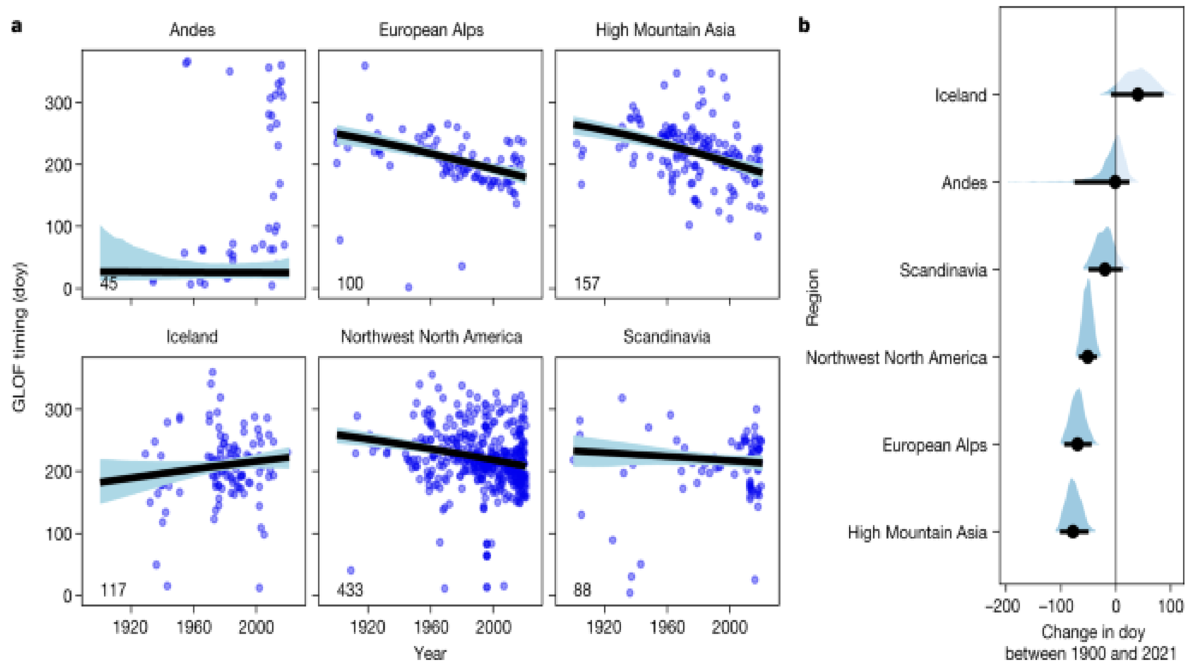
Recent outbursts of Lac de Faverges (Switzerland) may explain the only increase in the 10% highest records of Q_p in the European Alps since 1990 (Fig. 2), although reported values of Q_p in this region are very low (<100 $m^3 s^{-1}$) by global standards[26,27]. Regional trends in extreme V_0 since 1990 mirror those of extreme Q_p, and are negative in five of the six regions. Except for a negative trend in High Mountain Asia, all regional trends in median V_0 remain unchanged. By explicitly drawing on stratified subsets of the data, we conclude that regional trends in GLOF size over the recent shorter time interval with probably better documentation hardly deviate from the long-term trends. The recent decrease in extremes is striking because these arguably represent the best documented events with large values of Q_p and V_0, making a potential increase in extreme GLOFs with time very unlikely.

Earlier annual timing of GLOFs

Most GLOFs happen in summer, and winter remains largely free of GLOFs (Extended Data Fig. 6). In the Northern Hemisphere, GLOF activity peaks in July and August, and in the Southern Hemisphere in January when monthly average temperatures are highest, thus supplying ice-dammed lakes with large amounts of meltwater. Yet, the average date of outbursts within a given year has shifted earlier globally by about 40 days compared with 1900. The strongest regional shifts in timing are in High Mountain Asia (77_{-29}^{+25} days earlier), the European Alps (70_{-26}^{+24} days earlier) and northwest North America (52_{-19}^{+25} days earlier) (Fig. 3). In contrast, GLOFs in Iceland occur about six weeks later in the year on average, with the physical causes yet to be determined. At the site scale, we find that 68% of all ($n = 40$) ice-dammed lakes for which there is a documented day of the year (doy) burst out earlier in the year on average; in 30% of these cases, trends are credibly negative (Extended Data Fig. 7). The strongest change occurred at Lake Gornersee (Switzerland), which bursts out almost four months (114_{-19}^{+16} days) earlier today than at the beginning of the twentieth century[22]. Successively smaller and earlier floods could signal a local weakening, if not the end, of the GLOF cycle[21]. Indeed, already half

(46) of the 101 glaciers with two or more failures stopped releasing GLOFs before 2000. Occasionally, human intervention interrupts the natural GLOF cycle, especially in cases where outbursts have repeatedly damaged infrastructure. In the European Alps and in Scandinavia, engineered structures such as drainage tunnels through bedrock or glacier ice stopped the recurrent outbursts of at least six ice-dammed lakes[28,29].

Fig. 3: Changes in the annual timing of ice-dam failures between 1900 and 2021.



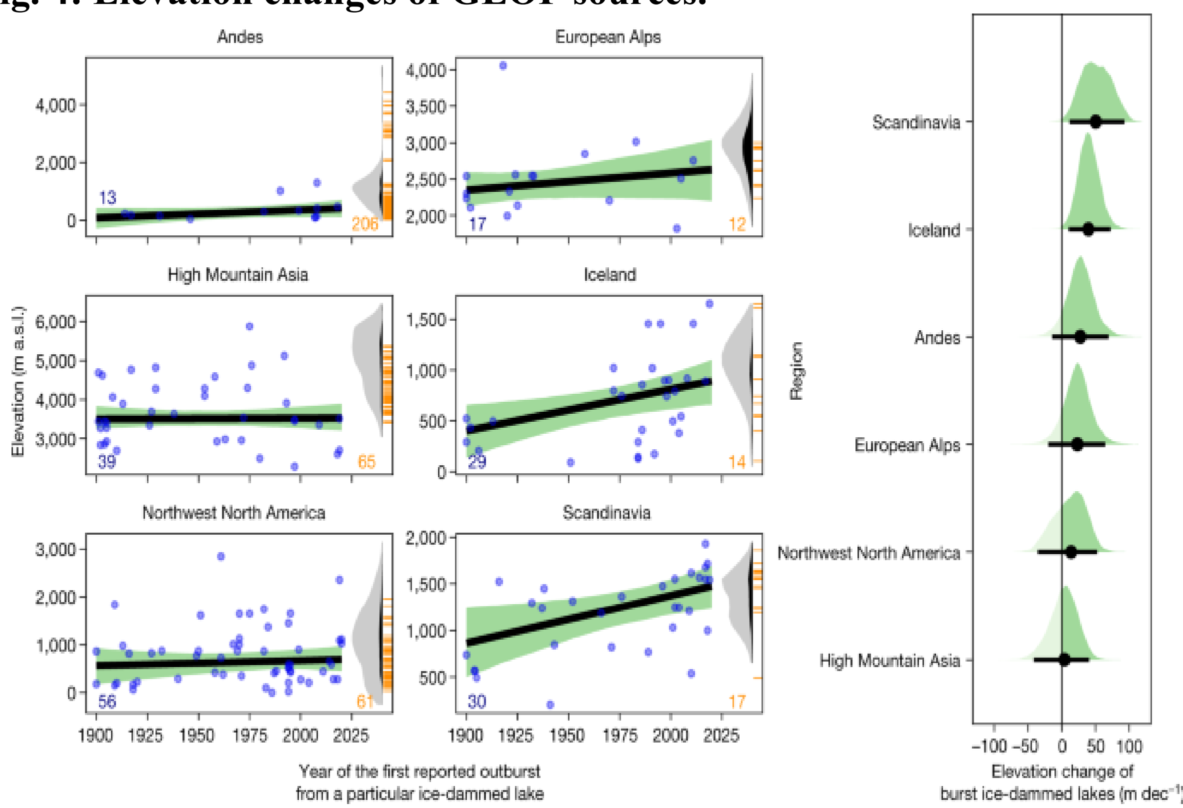
a, Regional trends in the annual date of GLOFs. Thick black lines and light blue shades show, respectively, the median and 95% HDI of changes in the doy when glacier lakes burst out. Numbers in lower left corners are sample sizes. **b**, Posterior differences in GLOF timing (doy) between 1900 and 2021. Black bubbles are medians and horizontal lines are 95% posterior HDIs. Probability masses of negative trends are dark.

Increasing elevation of ice-dam failures

Our database further shows that outbursts from ice-dammed lakes originated at progressively higher elevations over the past 12 decades (Fig. 4). The largest elevation shifts are in Scandinavia, Iceland and the Andes, where new sources of GLOFs moved 25–50 m higher each decade on average.

Outlet glaciers in these regions retreated rapidly in past decades, eliminating dams at their toes[25,30,31]. GLOFs in the European Alps had no credible change in elevation during our study period, although many large dams formed at lower elevations before 1900, when glaciers descended down to major valleys[2]. These dams disappeared with warming and glacier retreat after the mid-nineteenth century[2]. In contrast, High Mountain Asia has had more than 100 valley-blocking glacier advances to lower elevations in the past three decades[3,32], explaining the unchanged elevation trend of ice-dam failures in that region. In most regions, outbursts emerged from lakes that were tens to many hundreds of metres lower in elevation than today (Fig. 4 and Extended Data Fig. 8). In three regions, the number of ice-dammed lakes decreased over time: Scandinavia, the European Alps and Iceland have, in total, only 25 ice-dammed lakes >0.01 km^2 today, whereas three times as many lakes burst out in the past 120 yr (Fig. 4).

Fig. 4: Elevation changes of GLOF sources.



a, Changes in elevation of ice-dam failures over time. Blue bubbles mark years of the first reported outburst and the elevations above sea level (a.s.l.) of ice-dammed lakes. Thick black lines and green shades are the median and

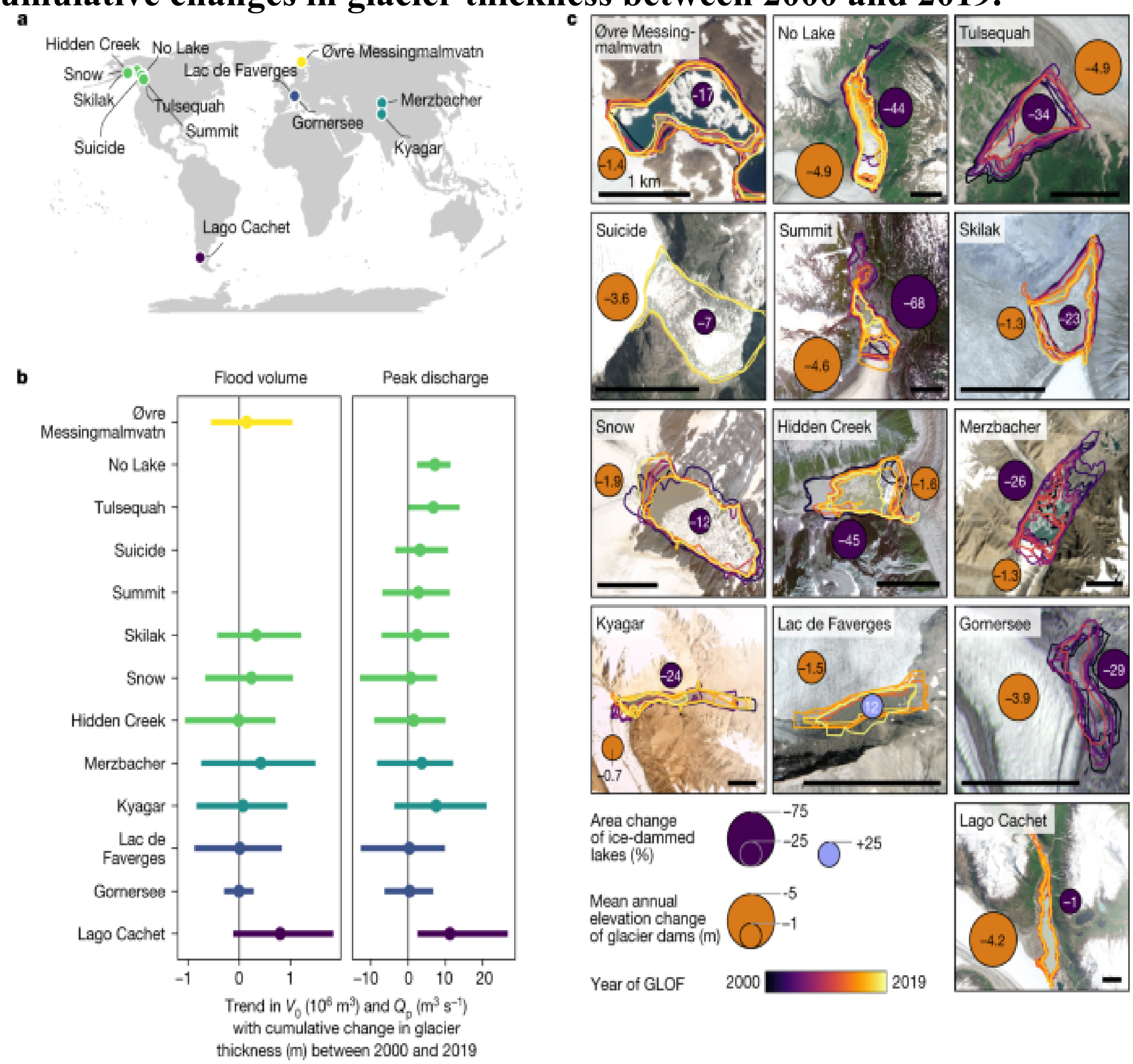
95% HDI of the posterior trends, respectively. Blue numbers in lower left corners are the number of lakes that burst out between 1900 and 2021. Orange ticks and numbers on the right are elevations and the number of present-day ice-dammed lakes >0.01 km² in each region, respectively[48,55,56,57,58]. The grey densities estimate present-day glacier volumes across elevation from ref. [47], truncated at 95% of the data. The black areas within the grey densities show the loss of glacier volume over the period 2000–2019 from ref. [8], relative to the entire glacier volume. **b**, Posterior decadal trends in annual elevation change of GLOF sources. Black bubbles are medians and horizontal lines are 95% posterior HDIs. Probability masses of positive trends are dark.

Impact of glacier decay on GLOF size

How does overall glacier decay help explain changes in ice-dam failures? In theory, the drainage of ice-dammed lakes occurs when the water pressure in the lake is close to or exceeds the pressure of the adjacent ice dam[33,34,35]. A GLOF may initiate once a critical ratio between lake depth and glacier thickness is reached[20]. As glaciers thin due to atmospheric warming[8], the necessary water depth to trigger a GLOF should lower, resulting in smaller lakes that produce floods of decreasing magnitudes[20]. We searched for a possible correlation between glacier thinning and GLOF magnitudes by focusing on 12 glaciers (one glacier impounds two lakes), for which we could obtain the cumulative elevation change of the ice dam, maximum lake areas before outbursts and reported GLOF magnitudes (V_0 or Q_p) between 2000 and 2019 (Methods). Over this period, we observed a thinning of the glacier dams by $\({36}_{-62}^{+19}\,{\rm{m}}\)$ (median and 95% HDI; Fig. 5c and Extended Data Fig. 9), and that 12 of 13 lakes shrank, with 8 losing more than 20% of their surface areas (Fig. 5c). Yet, rapid glacier thinning appears to correlate only moderately with reported GLOF magnitudes: average trends between glacier thinning and decreases in flood volumes or peak discharges are positive for only half of the glacier dams (Fig. 5a,b). Credibly positive trends occur only at Tulsequah Lake and Lake No Lake (British Columbia), and at Lago Cachet (Patagonia). Most other lakes have no clear trend. Interactions between ice and water might be an important mechanism for removing parts of glacier dams, causing most ice-

dammed lakes to shift laterally towards the glacier dam during the GLOF cycle. For example, Lake No Lake and Summit Lake (northwest North America) eroded several hundreds of metres from their ice dams in the past 20 yr (Fig. 5c).

Fig. 5: Trends in flood volume, peak discharge and lake area with cumulative changes in glacier thickness between 2000 and 2019.



a, Map of 15 ice-dammed lakes that produced more than five GLOFs each in that period. We generated the map in the free software RStudio V2022.07.2 (https://posit.co/products/open-source/rstudio/) using the ggmap package, including free data on administrative boundaries from the rnaturalearth package. **b**, Posterior regression slopes of V_0 (left) and Q_p (right) versus

cumulative change in thickness of the ice dam. Bubbles are medians and horizontal lines are 95% posterior HDIs of the posterior trends. Colours in **a** and **b** distinguish the study regions. Missing trends for individual lakes arise from lack of data. **c**, Local changes in lake area and glacier elevation between 2000 and 2019. We mapped lake areas (polygons) from satellite imagery immediately before the outburst using QGIS V3.22.6. Colours show the year of the GLOFs. Blue bubbles show the average percent change in lake area between the first and last reported GLOF for each lake. Orange bubbles show the mean annual elevation change of the glacier dam. Black horizontal scale bar is 1 km in all panels in **c**. Data on lake and glacier elevation change are available at ref. [27]. All background imagery are from 2019 Planet Labs Inc., obtained under a basic license in the Education and Research programme. We generated all maps in **c** using QGIS V3.22.6.

The unclear relationship between changes in glacier thickness and GLOF magnitudes suggests a nonlinear response, or instead that glacier thinning is a poor proxy for GLOF magnitudes. Variances among the estimated local trends point to local drainage mechanisms. Small variances in Q_p and V_0 (ref. [27]) may indicate recurrent opening and closing of the same subglacial tunnel system that allows the floods to attain high peak flows despite decreases in glacier thickness[22]. Floating ice can further exert high pressures on the underlying water in lakes, maintaining high Q_p and V_0 independent of glacier thinning that reduces the basin's storage capacity[36]. In contrast, high variances or abrupt changes in the trends may indicate shifts between drainage regimes, for example due to failure of ice buttresses at the point of inflow into the glacier[19]. Ice motion during an outburst can also close subglacial tunnels before the lake has emptied completely[34]. Some lakes such as Lago Cachet (Fig. 5c) have maintained their surface area but gained depth as sediment was removed from the lake bottom between successive outbursts[12]. Outburst floods can also drain other water bodies within and beneath the glacier. The volume from these sources can rival or even exceed those of the subaerial lake, although distinguishing the contributions from the individual sources is subject to high uncertainties[37,38]. Local weather can also add to the observed variability. Water supply to lakes can suddenly increase following heavy precipitation or during heatwaves when there is high glacier and snow melt, resulting in unusual high peak flows[16,39].

Accurately determining such local water fluxes could make GLOFs more predictable[22,37,40], although it would require a much denser meteorological network to quantify weather conditions in ice-dammed basins[41].

Implications for GLOF hazard

In our study regions, >70% of all reported GLOFs since 1900 originated from ice-dammed lakes[24]. However, their present-day number and sizes are small compared with meltwater lakes dammed by moraines or bedrock. In northwest North America, ice-dammed lakes account only for 6% of the total lake abundance[18], and in High Mountain Asia, as few as ~0.06% of total lake area is impounded by glaciers[42]. However, given their high outburst activity, we infer that ice-dammed lakes are among the most hazardous of all meltwater lakes. Still, catastrophic damage and fatalities remain a rare consequence of ice-dammed failures: only 10% of all floods in our database have documented socio-economic impacts, mostly involving destroyed bridges, pipelines, roads and campgrounds buried in debris, or loss of farmland and cattle affecting local food supply. Agricultural loss prevails in reports from Scandinavia, Iceland, the Andes and the European Alps. Since 1900, three ice dam failures have claimed human lives, although loss of life was reported more frequently in previous centuries. Well documented cases include the sudden failures of an ice-dammed lake in the Dranse Valley, Swiss Alps, in 1590 and 1818 that killed at least 40 and 140 people, respectively[43]. Such one-time ice-dam failures and those starting a new, unexpected GLOF cycle with flood defence measures not yet in place have caused much damage[16,44]. Dykes, floodgates, dams, or the temporary or permanent resettlement of mountain communities have reduced the impact of recurrent ice-dam failures, even for exceptionally large floods[1].

The global reduction in glacier volumes[8] will probably control ice-dam failure in the future. We find that 90% of all outburst floods in our database originated below the elevation band where most glacier mass is stored today (Fig. 4a). Elevations of historical ice-dam failures are largely within glacier ablation areas, where glaciers melt fastest and temporarily trap excess meltwater at their margins. This zone has been rapidly shifting upwards under ongoing atmospheric warming. In the European Alps and the

Himalayas, for example, glaciers have been melting throughout their entire elevation range in the twenty-first century[45,46], and ice-dammed lakes are now forming even in the highest parts of ice-covered catchments (Fig. 4a and Extended Data Fig. 8). Progressive retreat at the glacier terminus may thus push the accommodation space for new lakes towards the natural limit set by the remaining glacier mass close to the drainage divide. According to projections, regions with small amounts and volumes of glaciers, such as the European Alps, Scandinavia and British Columbia, could become largely ice free by 2100[9]. We anticipate that the window for future ice-dam formation and failure might close within a few decades in these mountain ranges. Glacier remnants in steeper and higher terrain may be capable of only damming, if at all[5], smaller lakes, albeit with higher potential energy. In the course of one to several centuries, ice-dam failures may diminish as a cryospheric hazard, although mass flows from other lakes, possibly tied to permafrost thaw, may take over.

This trajectory of changing GLOF hazard is likely to differ in regions with large glacier volumes. Alaska and Patagonia, for example, host ~70% of the total glacier volume in our study regions, about half of which will persist beyond the end of the twenty-first century[9,47]. Extensive glacier streams currently terminate at or near sea level, and hundreds of glacier-dammed lakes have been forming in ice-free tributary valleys in recent decades[18,48] (Fig. 4a). Yet, statistical models have failed to identify glaciological or meteorological predictors that could explain the growth pattern of ice-dammed lakes in northwest North America[49]. Trends in the magnitude and timing of GLOFs, let alone any drivers, are largely unknown for most of these lakes[23]. Newly exposed proglacial areas could be sources of new cycles of GLOFs, posing hazards for potential development in tourism, mining and hydropower generation in these areas. In this regard, time series of satellite images[50] and stream gauges[12,14,22] are invaluable archives to systematically capture geomorphic changes caused by GLOFs. Integrating newly detected and reported cases into a single database is thus key to quantifying and predicting climate-driven changes in GLOF magnitude, frequency and hazard[24].

In summary, our analysis reveals a strong decay in the highest GLOF magnitudes and discharges, whereas GLOFs of median size have changed

little in the past 12 decades. Given their steep hydrographs, floods caused by ice-dam failures will remain discharge extremes, even during glacier retreat. How and where the remaining glaciers will form dams in the future will determine the extent to which GLOFs will feed the sediment cascade[12], disturb ecological and geochemical cycles in mountain rivers[51,52], and locally inundate floodplains and terraces[7]. To anticipate these consequences, future work will be required to refine the few available physical models that link glacier retreat and lake formation with the size and timing of outbursts[10,34,35,36]. Recently released high-resolution data on elevation, glacier change, ice-flow velocities and climate will be key to understanding better the nexus between meltwater accumulation and hazard. Predicting and monitoring the locations and dynamics of future lakes will become increasingly important, especially in steep mountain regions that are difficult to access, but have growing populations and infrastructure downstream. International collaborations such as the United Nations Development Programme project ‘GLOF-I/-II’ in Pakistan aim to mitigate GLOF hazards, but more effort will be required to manage the episodic GLOF discharges as the variability in precipitation, temperature and river discharge rises[53,54]. Our dataset forms a solid base to help predict changes in GLOF hazard from the catchment to the regional scale to ensure more sustainable development for one of the most sensitive regions on Earth.

Methods

Developing a database of ice-dam failures

We define an ice-dammed lake as any water body at the margin of and impounded by glacier ice. We exclude all other glacier-fed lakes, such as lakes dammed by moraines or landslides, internal water pockets, supraglacial ponds, lakes below glaciers and lakes formed by geothermal activity. Ice dams often trap runoff from catchments and cause loss and damage when the impounded water is suddenly released and reaches human settlements[59]. We focus on six major glaciated mountain regions that support 99% of the contemporary ice volume outside polar regions[47]. Our evidence for the 1,569 ice dam failures reported between 1900 and 2021 comes from a total of 446 different sources of information. We distinguish between

primary sources, where we have direct observations or access to raw data, and secondary sources that describe, interpret or summarize previous findings on historical GLOFs. Primary sources are available for 64% of all reported cases and include original scientific publications (36%), annals and reports published by local and regional authorities (18%), written communications in emails with eyewitnesses and experts (9%), and newspapers and internet resources such as video footage and posts on social media platforms (altogether 1%). Secondary sources (36% of all cases) include second-hand information derived from reviews and summaries of previous publications (19%) and entries in databases with no direct access to the primary source (17%). Half of all sources report only a single event, whereas a few studies include information on dozens of GLOFs. For example, ref. [60] identified more than 100 outbursts from lakes impounded by Kennicott Glacier during the twentieth century by studying stream gauge data and historical aerial photographs. Still, decades may elapse between the date of a GLOF and the reporting, especially where researchers reprocessed and reanalysed historical data (Extended Data Fig. 1). In half of all cases, however, GLOFs were reported within a decade after they occurred. A few studies contributed significantly to improving the regional record of GLOFs in the first decades of the twentieth century, when the overall reporting rate was low[3,7,11,60,61,62]. In a few cases, reporting may have changed due to resettlement in this period. For example, repeated flooding of an ice-dammed lake below the Folgefonni Ice Cap (Norway) forced farmers to abandon their homes in the mid-1960s[16]. It is only since the 1970s that research activity on GLOFs has become more consistent[24]: 63% GLOFs in our databases were first documented in that period (Extended Data Fig. 1).

In collecting historic GLOFs, we only considered cases with either a documented date (at least the year of occurrence) or a time interval of failure, centroidal coordinates of the lake and at least one reference. For each GLOF, we documented the mountain range, the country from which the flood originated, the name of the parent glacier (local name and the ID in the Randolph Glacier Inventory RGI V6.0[63]) and the name of the burst glacier lake. We extracted the surface elevation of each lake centroid from the Advanced Land Observing Satellite (ALOS) V3.2 digital elevation model (DEM)[64]. We collected the reported flood volume ($n = 510$, 33% of all

cases), peak discharge ($n = 506$, 33%) and timing ($n = 940$, 61%). Finally, we documented any impacts, damage, and losses for events in our database.

Lake mapping

We could verify the origin, timing and area of 53% of all reported lake outbursts that have occurred since the Landsat 5 mission began to continuously obtain satellite images in the late-1980s[65]. Later generations of satellites, such as RapidEye and the Planet Cubesat constellation (launched in 2009 and 2015, respectively), have higher spatial resolution and thus capture GLOF features in greater detail. Ice-dammed lakes decrease in size, or even empty entirely, during GLOFs, exposing either partly or fully the lake floor, which is often covered with stranded icebergs (Fig. 5c). Ice, clouds, and shadows compromise the automatic detection of water and lakes[66]. Hence, we manually digitized the extents of the lakes from satellite images in QGIS V3.16 software. To this end, we scanned the archive of satellite images to retrieve the last available image before and the first available image after the lake drained. Mapping lakes from optical images was limited in Iceland because of frequent cloud cover. For each lake with repeat outbursts, we used a simple linear regression model of the mapped lake area before the outbursts versus time to obtain the change in lake area (Fig. 5). We then estimated the mean percent change in lake area between the first and last reported outburst in the time series. For manually mapped lake areas, uncertainties are generally assumed to be between 4% and 10% of the mapped lake area[48,55] or between 0.5 and 1 pixel of the sensor resolution[18,42,57], that is, 3–5 m for Planet and RapidEye, and 30 m for Landsat images.

In Fig. 4, we compare elevations of lakes that have had historic outbursts with elevations of ice-dammed lakes that currently (2018–2020) exist according to regional inventories[18,48,55,56,57,58]. The lakes in these databases were either manually digitized or refined after automatic mapping. We chose a minimum mapping unit of 0.01 km^2 to ensure comparability between inventories. No contemporary lake inventory is available for Iceland. Thus, we manually mapped ice-dammed Icelandic lakes from high-resolution Google Earth imagery obtained between May and September 2021. We

added the elevations obtained from the ALOS V3.2 DEM[64] to both the previous lake inventories and our manually mapped lakes.

Bayesian hierarchical modelling

To assess temporal changes in GLOF magnitude, elevation and timing, we fitted Bayesian hierarchical models that distinguish between regional (that is, our six study regions) or local (that is, lakes with repeat outbursts) groups. Bayesian models form a compromise between the likelihood of observing the data under the assumption of specific model parameters and prior knowledge about these parameters. We encode our uncertainty in this prior knowledge before obtaining a posterior distribution for all parameters conditioned on the data. The benefit of hierarchical models is that they account for group-level structure in the data within a single model. We can thus estimate regional or local effects with respect to a population-level model learned from all data regardless of their location. Hierarchical models offer a compromise between separate models that overemphasize regional and local effects and that are independently learned for each group, and pooled models that are learned from all data without distinction[67]. Hierarchical models improve parameter estimates for individual groups, especially if sample sizes are low or vary across the groups. This trait is particularly desirable for our study as the number of reported GLOFs differs widely among both regions and lakes. The model learns the population-level parameters from the data, and these parameters are shared among groups. In this way, the groups inform each other, achieving an effective sample size that greatly exceeds that of single groups. The joint posterior density in a hierarchical model is proportional to the product of the likelihood and the joint prior density $\pi ({\theta }_{1},\ldots ,{\theta }_{J},\tau |D)\propto$ $\pi (D|{\theta }_{1},\ldots ,{\theta }_{J},\tau)\pi ({\theta }_{1},\ldots ,{\theta }_{J},\tau)=\pi (\tau){\prod }_{j=1}^{J}\pi ({D}_{j}|{\theta }_{j})\pi ({\theta }_{j}|\tau)$, where D are the observed data and D_j is the subset of D that contains all data in group j. The group-level parameters ${\theta }_{j}=({\theta }_{1j},\ldots ,{\theta }_{Kj})$ define the model on the group-level with K individual parameters, which vary for $j=1,\ldots ,J$ groups. These parameters are drawn from distributions $\pi \left({\theta }_{j}|\tau \right)$ and specified by population-level (hyper-) parameters $\tau =\left({\tau }_{1},\ldots ,{\tau }_{L}\right)$, which consist of L

individual parameters and are also learned from the data. The hierarchical model estimates the posterior distribution for both population-level and group-level parameters conditional on the reported GLOFs.

Assessing regional and local trends of V_0, Q_p and Z

Continental and global studies of river floods point to climate-driven changes in both median and extreme discharges in past decades[68,69,70,71]. Similarly, repeated inventories reveal recent shifts in the regional size distribution of ice-dammed lakes, but little about trends in lake outbursts[18,48,49]. By using quantile regression, we examined whether and by how much the quantiles in the conditional distributions of reported peak discharge, Q_p, and flood volume, V_0, have changed in our six study regions. Extreme (>90th percentile) flood discharges are important for planning flood protection measures and informing flood risk zoning[69]. Ordinary least squares regression estimates the conditional mean, but quantile regression makes no assumptions about the distribution of the response variable and is robust against outliers. In our model specification, we used an asymmetric Laplace likelihood (AL) with fixed quantiles. The AL has density

$${f}_{{\rm{AL}}}(\,y{\rm{| }}\mu ,\kappa ,p)=\frac{p(1-p)}{\kappa }\exp \left(-{\rho }_{p}\left(\frac{y-\mu }{\kappa }\right)\right),$$
(1)

where y is the response variable, μ is a location parameter, κ is a positive scale parameter, $0 < p < 1$ is a percentile, and ${\rho }_{p}(x)=x(p-I(x < 0))$ with indicator function $I(\cdot)$. The exponent takes the values $-\frac{\,|{y}-\mu |}{\kappa }p$ for $y \geq \mu$ and $-\frac{|y-\mu \,|}{\kappa }(1-p)$ for $y < \mu$. We modelled the conditional distributions of Q_p and V_0 with decimal year t for two fixed values of p as

$${y}_{ji}\sim {\rm{A}}{\rm{L}}({\mu }_{ji},\kappa ,p),{\rm{f}}{\rm{o}}{\rm{r}}\,j=1,\ldots ,\,J\,{\rm{a}}{\rm{n}}{\rm{d}}\,i=1,\ldots ,{n}_{j}$$
(2)

$${{\rm{\mu }}}_{ji}={\alpha }_{j}+{\beta }_{j}{t}_{ji},{\rm{for}}\,j=1,\ldots ,\,J\,{\rm{and}}\,i=1,\ldots ,{n}_{j}$$

(3)

$$[\begin{array}{c}{\alpha }_{j}\\ {\beta }_{j}\end{array}]\sim {\rm{M}}{\rm{V}}{\rm{N}}{\rm{o}}{\rm{r}}{\rm{m}}{\rm{a}}{\rm{l}}\,[(\begin{array}{c}\alpha \\ \beta \end{array}),S],\,{\rm{f}}{\rm{o}}{\rm{r}}\,j=1,\ldots ,J$$

(4)

$$S=\left(\begin{array}{cc}{\sigma }_{\alpha } & 0\\ 0 & {\sigma }_{\beta }\end{array}\right)R\left(\begin{array}{cc}{\sigma }_{\alpha } & 0\\ 0 & {\sigma }_{\beta }\end{array}\right)$$

(5)

$$R=\left(\begin{array}{cc}1 & \varsigma \\ \varsigma & 1\end{array}\right)$$

(6)

where y_{ji} are reported values of Q_{p} or V_0 with y_{ji} referring to the ith observation in group j, n_j is the number of observations in group j, and J is the number of groups. The parameters α_j and β_j are the group-level intercepts and slopes, respectively, and α and β are the corresponding population-level parameters. The covariance matrix S is composed of group-level standard deviations σ_α and σ_β, and R, the correlation matrix with correlation ς. In line with decadal analyses of median and extreme flood discharges[68,69,70], we considered fixed values of $p = 0.5$ (median) and $p = 0.9$ (90th percentile), and estimated all other parameters from the data.

The Bayesian framework demands prior distributions for each parameter that enters at another model level. We chose the following priors

$$\kappa \sim N(0,2.5)$$

(7)

$$\alpha \sim N(0,2.5)$$

(8)

$$\beta \sim N(0,2.5)$$

(9)

$${\sigma }_{\alpha }\sim N(0,2.5)$$

(10)

$${\sigma }_{\beta }\sim N(0,2.5)$$

(11)

$$R\sim {\rm{L}}{\rm{K}}{\rm{J}}{\rm{C}}{\rm{h}}{\rm{o}}{\rm{l}}{\rm{e}}{\rm{s}}{\rm{k}}{\rm{y}}(1).$$
(12)

These priors refer to standardized, $\log_{10}$-transformed response variables $Q_{\rm p}$ and V_0, and a standardized predictor (decimal year of observation) with zero mean and unit standard deviation. Our choice of a zero-mean Gaussian with standard deviation of 2.5 admits both negative and positive trends for β, expressing the contrasting trends reported for the size of ice-dam failures in recent decades[1,6,19,22,72]. The Lewandowski–Kurowicka–Joe (LKJ) Cholesky correlation distribution prior for R makes all correlation matrices equally likely.

For a fixed quantile, the joint posterior distribution can be thus written as

$$\begin{array}{c}\pi ({\alpha }_{1},...,{\alpha }_{J},{\beta }_{1},...,{\beta }_{J},\kappa ,\alpha ,\beta ,{\sigma }_{\alpha },{\sigma }_{\beta },R|y,t,p)\propto \pi (\kappa ,\alpha ,\beta ,{\sigma }_{\alpha },{\sigma }_{\beta },R)\\ \,({\prod }_{j=1}^{J}{\prod }_{i=1}^{{n}_{j}}{f}_{{\rm{A}}{\rm{L}}}(\,{y}_{ji}|{\alpha }_{j}+{\beta }_{j}{t}_{ji},p,\kappa)\pi ({\alpha }_{j},{\beta }_{j}|\alpha ,\beta ,{\sigma }_{\alpha },{\sigma }_{\beta },R)),\end{array}$$
(13)

where $\pi \left({\alpha }_{j},{\beta }_{j}| \alpha ,\beta ,{\sigma }_{\alpha },{\sigma }_{\beta },R\right)$ is the prior distribution for the group-level parameters as defined in equations (4)–(6) and $\pi \left(\kappa ,\alpha ,\beta ,{\sigma }_{\alpha },{\sigma }_{\beta },R\right)$ is the joint prior distribution for the independent population-level parameters as defined in equations (7)–(12). We numerically approximate the posterior distribution using a Hamiltonian sampling algorithm implemented in Stan[73] that is called via the software package brms[74] within the statistical programming language R[75].

We used the same multi-level structure of the quantile regression model to determine how flood discharges have changed for individual lakes. To this end, we conditioned the model on $J = 22$ lakes ($Q_{\rm p}$) and $J = 26$ lakes (V_0)

that produced at least five outbursts each between 1900 and 2021. For these models, we focused only on the trends in median flood discharges, given that lakes with few reported GLOFs have few samples that exceed high thresholds.

Finally, to test whether more recent GLOFs originated from progressively higher elevations, we selected for each lake the year of the first reported GLOF, given that many ice-dammed lakes burst out repeatedly. As for our hierarchical models on GLOF magnitudes, we used quantile regression to determine the trend in the median elevation Z of glacier-dammed lakes against the decimal year of GLOF occurrence t, conditioned on the $J = 6$ regions.

We ran all regional and local quantile regression models on Q_p, V_0 and Z with three parallel chains that drew 6,000 samples with 2,000 warm-up runs each. We found no numerical divergences after warm-ups, which were supported by the Gelman–Rubin potential scale reduction factor \(\hat{R}=1.0\) in all models, indicating that the Markov chains have converged. We report the posterior distributions of the pooled model parameters in Supplementary Tables 1–3, and the group-level regression slopes in Extended Data Figs. 2, 3 and 5. These regression slopes are conventionally deemed ‘credibly’ different from zero if a select interval of the posterior probability mass, in our case the 95% HDI, excludes zero. This choice of the HDI is arbitrary, and we also report how much of the probability mass of the posterior distribution is on either side of zero (Extended Data Figs. 2–5).

Assessing trends in doy

Measuring temporal changes in GLOF timing must consider that the response variable doy can recycle through the observation period, given that calendar day 0 is equal to day 365. The von Mises (VM) distribution is a close approximation to a wrapped normal distribution for circular data and allows estimating the angular response d (doy) from the linear predictor year t (ref. [64]). The VM distribution has density

$${f}_{{\rm{VM}}}(\,y| {\vartheta },\phi)=\frac{{\rm{\exp }}(\phi \,\cos (\,y-{\vartheta }))}{2\pi {I}_{0}(\phi)},$$
(14)

where y are the observed data, ϑ is a location parameter, φ is a positive precision parameter and I_0 is the modified Bessel function of the first kind of order 0.

Following the parameterization in brms, we rescaled the response d to ${y}_{i}\in (-\pi ,\pi)$, standardized t to a zero mean and unit standard deviation, and estimated the trend in GLOF timing as

$${y}_{ji}\sim {\rm{V}}{\rm{M}}({\vartheta }_{ji},{\phi }_{j}),\,{\rm{f}}{\rm{o}}{\rm{r}}\,j=1,\,\ldots ,\,J\,{\rm{a}}{\rm{n}}{\rm{d}}\,i=1,\,\ldots ,\,{n}_{j}$$
(15)

$${{\vartheta }}_{ji}={\zeta }_{j}+{\eta }_{j}{t}_{ji},\,{\rm{for}}\,j=1,\,\ldots ,\,J\,{\rm{and}}\,i=1,\,\ldots ,\,{n}_{j}$$
(16)

$${\zeta }_{j}\sim N(0,2.5)$$
(17)

$${\eta }_{j}\sim N(0,2.5)$$
(18)

$${\phi }_{j}\sim {\rm{G}}{\rm{a}}{\rm{m}}{\rm{m}}{\rm{a}}(2,0.01),$$
(19)

where y_{ji} are reported values of doy, with y_{ji} referring to the ith observation in group j, ϑ_{ji} is the corresponding location parameter, and ζ_j and η_j are the estimated intercepts and slopes, respectively. We used the default Gamma prior in brms to estimate the precision parameter φ_j. This model was fit to each group separately without partial pooling of parameters across groups, because the distinct seasonal differences in GLOF timing in the two hemispheres (Extended Data Fig. 6) caused numerical divergences. Using a VM likelihood function, the joint posterior distribution can be written as

$$\begin{array}{c}\pi ({\zeta }_{j},{\eta }_{j},{\phi }_{j}|{y}_{j},{t}_{j})\propto ({\prod }_{i=1}^{{n}_{j}}{f}_{{\rm{V}}{\rm{M}}}({y}_{ji}|{\zeta }_{j}+{\eta }_{j}{t}_{ji},{\phi }_{j}))\,\pi ({\zeta }_{j},{\eta }_{j},{\phi }_{j}),\\ \,\,\,{\rm{f}}{\rm{o}}{\rm{r}}\,j=1,\,\ldots ,\,J\end{array}$$
(20)

where y_j and t_j are the observations and corresponding years of group j and $\pi ({\zeta }_{j},{\eta }_{j},{\phi }_{j})$ is the joint prior of the regression parameters. We maintained the set-up of the Hamiltonian sampler in brms, using three parallel chains each with 6,000 samples after 2,000 warm-up runs. We found that $\hat{R}=1.0$ in all runs, indicating convergent Markov chains. We report the posterior distributions of all model parameters in Supplementary Table 4.

Estimating local trends between glacier elevation change and V_0 and Q_p

The length and thickness of the glacier dam are important diagnostics for predicting flood magnitudes and timing[76]. Changes in dam geometry could therefore alter the pattern of lake outbursts, in both frequency and magnitude. In QGIS, we manually trimmed the area of all glaciers in the RGI V6.0 that impound a lake to the area below the dam. We define a dam as the area between the glacier terminus and the elevation contour on the glacier directly upstream of the lake. We then extracted the cumulative elevation change of all glacier dams between 2000 and 2019, following the procedure by ref. [8]. In summary, this method first generates DEMs from all available stereo-pairs of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) satellite instrument intersecting glaciers during the period of 2000 to 2019. We stacked the ASTER DEMs in temporal order and further added high-resolution data from the ArcticDEM[77] to the time series. Data from ArcticDEM cover glacier dams in Alaska, Iceland and Scandinavia (Supplementary Table 6). Each pixel in the DEM time series is then interpolated using a Gaussian process regression model to estimate the elevation at any date within this period. We extracted the elevation time series for the areas of the glacier dams, resulting in an area-weighted average of 45 independent DEMs between 2000 and 2019 for

each dam. Individual glacier dams are covered by at least 95% of measured pixels, drawing on at least 13 independent DEMs per pixel. We calculated the associated elevation differences between successive time steps, interpolated spatial data gaps hypsometrically[78] and finally aggregated them to the cumulative mean elevation changes with respect to 1 January 2000. On average, our glacier dams thin at a rate of 3 m yr^{-1} during the period 2000–2019, with an average uncertainty of ±0.23 m yr^{-1} (95% confidence level). The elevation change uncertainties were estimated using a framework that accounts for heteroscedasticity (that is, variability in precision), which more reliably describes errors depending on terrain slope and the quality of stereo-correlation. The framework also accounts for spatial correlation of errors using multiple correlation ranges that better describe ASTER instrument noise[79]. These uncertainties were extensively validated using independent high-precision data from NASA's Ice, Cloud and land Elevation Satellite (ICESat) and IceBridge missions, as well as LiDAR acquisitions[8]. Supplementary Table 6 reports statistics on spatial coverage, the temporal density of DEM data, thinning rates between 2000 and 2019, and associated uncertainties for each glacier dam.

To assess the impact of glacier thinning on GLOF magnitudes, we selected only $J = 13$ glacier lakes that burst out repeatedly ($n > 5$) between 2000 and 2019. For each dated outburst within a GLOF cycle, we obtained the closest temporal estimate of the cumulative elevation change of the glacier dam. We then fitted a hierarchical quantile regression model, that is, the same as in equations (1)–(13), to estimate the median of reported V_0 and Q_p at a given glacier dam from the associated cumulative elevation change h. As we did when calculating temporal trends of V_0, Q_p and Z, we chose an AL likelihood to estimate the conditional median of V_0 and Q_p, with the same (hyper-)prior distributions as in equations (7)–(12). We ran three parallel chains with 6,000 samples and 2,000 warm-up runs and found no numerical divergences. We report the posterior distributions of all model parameters in Supplementary Table 5.

Data availability

The GLOF inventory V2.0 is freely available at http://glofs.geoecology.uni-potsdam.de, and archived at https://doi.org/10.5281/zenodo.7326571. The ALOS DEM V3.2 is freely available at https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm. The RGI V6.0 is freely available at https://nsidc.org/data/nsidc-0770. Planet and RapidEye images are available under a basic license from the Education and Research programme from Planet Labs Inc. through the PlanetExplorer (https://www.planet.com/). Landsat images were obtained from the EarthExplorer (https://earthexplorer.usgs.gov/). Data on glacier elevation change for the period 2000–2019 are freely available at https://doi.org/10.6096/13.

Code availability

All codes to reproduce the statistical analysis are available at https://github.com/geveh/IceDamFailures, and are archived at https://doi.org/10.5281/zenodo.7326865.

References

1. Carrivick, J. L. & Tweed, F. S. A global assessment of the societal impacts of glacier outburst floods. *Glob. Planet. Change* **144**, 1–16 (2016).

2. Rabot, C. Glacial reservoirs and their outbursts. *Geogr. J.* **25**, 534–548 (1905).

3. Bazai, N. A. et al. Increasing glacial lake outburst flood hazard in response to surge glaciers in the Karakoram. *Earth Sci. Rev.* **212**, 103432 (2021).

4. Haeberli, W. Frequency and characteristics of glacier floods in the Swiss Alps. *Ann. Glaciol.* **4**, 85–90 (1983).

5. Liestøl, O. Glacier dammed lakes in Norway. *Nor. Geogr. Tidsskr.* **15**, 122–149 (1956).

6. Thorarinsson, S. Chapter IX. The ice dammed lakes of Iceland with particular reference to their values as indicators of glacier oscillations. *Geogr. Anna.* **21**, 216–242 (1939).

7. Hewitt, K. & Liu, J. Ice-dammed lakes and outburst floods, Karakoram Himalaya: historical perspectives on emerging threats. *Phys. Geogr.* **31**, 528–551 (2010).

8. Hugonnet, R. et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature* **592**, 726–731 (2021).

9. Hock, R. et al. GlacierMIP—a model intercomparison of global-scale glacier mass-balance models and projections. *J. Glaciol.* **65**, 453–467 (2019).

10. Roberts, M. J. Jökulhlaups: a reassessment of floodwater flow through glaciers. *Rev. Geophys.* **43**, 21 (2005).

11. Post, A. & Mayo, L. R. *Glacier Dammed Lakes and Outburst Floods in Alaska.* Hydrologic Atlas 455 (US Geological Survey, 1971); https://doi.org/10.3133/ha455.

12. Jacquet, J. et al. Hydrologic and geomorphic changes resulting from episodic glacial lake outburst floods: Rio Colonia, Patagonia, Chile. *Geophys. Res. Lett.* **44**, 854–864 (2017).

13. Mathews, W. H. & Clague, J. J. The record of jökulhlaups from Summit Lake, northwestern British Columbia. *Can. J. Earth Sci.* **30**, 499–508 (1993).

14. Geertsema, M. & Clague, J. J. Jökulhlaups at Tulsequah Glacier, northwestern British Columbia, Canada. *Holocene* **15**, 310–316 (2005).

15. Larsen, I. J. & Lamb, M. P. Progressive incision of the Channeled Scablands by outburst floods. *Nature* **538**, 229–232 (2016).

16. Tvede, A. M. Floods caused by a glacier-dammed lake at the Folgefonni Ice Cap, Norway. *Ann. Glaciol.* **13**, 262–264 (1989).

17. Wang, Y., Wu, N., Kunze, C., Long, R. & Perlik, M. in *The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People* (eds Wester, P. et al.) 17–56 (Springer, 2019); https://doi.org/10.1007/978-3-319-92288-1_2.

18. Rick, B., McGrath, D., Armstrong, W. & McCoy, S. W. Dam type and lake location characterize ice-marginal lake area change in Alaska and NW Canada between 1984 and 2019. *Cryosphere* **16**, 297–314 (2022).

19. Roberts, M. J., Pálsson, F., Gudmundsson, M. T., Björnsson, H. & Tweed, F. S. Ice–water interactions during floods from Grænalón glacier-dammed lake, Iceland. *Ann. Glaciol.* **40**, 133–138 (2005).

20. Tweed, F. S. & Russell, A. J. Controls on the formation and sudden drainage of glacier-impounded lakes: implications for Jökulhlaup characteristics. *Prog. Phys. Geogr.* **23**, 79–110 (1999).

21. Evans, S. G. & Clague, J. J. Recent climatic change and catastrophic geomorphic processes in mountain environments. *Geomorphology* **10**, 107–128 (1994).

22. Huss, M., Bauder, A., Werder, M., Funk, M. & Hock, R. Glacier-dammed lake outburst events of Gornersee, Switzerland. *J. Glaciol.* **53**, 189–200 (2007).

23. Capps, D. M. & Clague, J. J. Evolution of glacier-dammed lakes through space and time; Brady Glacier, Alaska, USA. *Geomorphology* **210**, 59–70 (2014).

24. Veh, G. et al. Trends, breaks, and biases in the frequency of reported glacier lake outburst floods. *Earths Future* **10**, e2021EF002426 (2022).

25. Dussaillant, I. et al. Two decades of glacier mass loss along the Andes. *Nat. Geosci.* **12**, 802–808 (2019).

26. Raymond, M., Wegmann, M. & Funk, M. *Inventar Gefährlicher Gletscher in der Schweiz.* Report 182 (Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie der Eidgenössischen Technischen Hochschule Zürich, 2003).

27. Veh, G. Supplementary data for 'Less extreme and earlier outbursts of ice-dammed lakes since 1900'. *Zenodo* https://doi.org/10.5281/zenodo.7326571 (2022).

28. Haeberli, W., Kääb, A., Mühll, D. V. & Teysseire, P. Prevention of outburst floods from periglacial lakes at Grubengletscher, Valais, Swiss Alps. *J. Glaciol.* **47**, 111–122 (2001).

29. Jackson, M. & Ragulina, G. *Inventory of Glacier-Related Hazardous Events in Norway.*Report 83 (Norwegian Water Resources and Energy Directorate, 2014).

30. Guðmundsson, S. et al. Terminus lakes on the south side of Vatnajökull Ice Cap. *SE Iceland. Jökull* **69**, 1–34 (2020).

31. Andreassen, L. M., Elvehøy, H., Kjøllmoen, B. & Belart, J. M. C. Glacier change in Norway since the 1960s—an overview of mass balance, area, length and surface elevation changes. *J. Glaciol.* **66**, 313–328 (2020).

32. Vale, A. B., Arnold, N. S., Rees, W. G. & Lea, J. M. Remote detection of surge-related glacier terminus change across High Mountain Asia. *Remote Sens.* **13**, 1309 (2021).

33. Röthlisberger, H. Water pressure in intra- and subglacial channels. *J. Glaciol.* **11**, 177–203 (1972).

34. Nye, J. F. Water flow in glaciers: Jökulhlaups, tunnels and veins. *J. Glaciol.* **17**, 181–207 (1976).

35. Fowler, A. C. Breaking the seal at Grímsvötn, Iceland. *J. Glaciol.* **45**, 506–516 (1999).

36. Jenson, A., Amundson, J. M., Kingslake, J. & Hood, E. Long-period variability in ice-dammed glacier outburst floods due to evolving catchment geometry. *Cryosphere* **16**, 333–347 (2022).

37. Bigelow, D. G. et al. The role of englacial hydrology in the filling and drainage of an ice-dammed lake, Kaskawulsh Glacier, Yukon, Canada.

J. Geophys. Res. Earth Surf. **125**, e2019JF005110 (2020).

38. Shangguan, D. et al. Quick release of internal water storage in a glacier leads to underestimation of the hazard potential of glacial lake outburst floods from Lake Merzbacher in central Tian Shan Mountains. *Geophys. Res. Lett.* **44**, 9786–9795 (2017).

39. Ng, F., Liu, S., Mavlyudov, B. & Wang, Y. Climatic control on the peak discharge of glacier outburst floods. *Geophys. Res. Lett.* **34**, L21503 (2007).

40. Kingslake, J. & Ng, F. Quantifying the predictability of the timing of jökulhlaups from Merzbacher Lake, Kyrgyzstan. *J. Glaciol.* **59**, 805–818 (2013).

41. Kienholz, C. et al. Deglacierization of a marginal basin and implications for outburst floods, Mendenhall Glacier, Alaska. *Front. Earth Sci.* **8**, 137 (2020).

42. Maharjan, S. B. et al. *The Status of Glacial Lakes in the Hindu Kush Himalaya* Research Report 2018/1 (ICIMOD, 2018).

43. Ancey, C. et al. Hydraulic reconstruction of the 1818 Giétro glacial lake outburst flood. *Water Resour. Res.* **55**, 8840–8863 (2019).

44. Mathews, W. H. Two self-dumping ice-dammed Lakes in British Columbia. *Geogr. Rev.* **55**, 46–52 (1965).

45. Fischer, M., Huss, M. & Hoelzle, M. Surface elevation and mass changes of all Swiss glaciers 1980–2010. *Cryosphere* **9**, 525–540 (2015).

46. Maurer, J. M., Schaefer, J. M., Rupper, S. & Corley, A. Acceleration of ice loss across the Himalayas over the past 40 years. *Sci. Adv.* **5**, eaav7266 (2019).

47. Farinotti, D. et al. A consensus estimate for the ice thickness distribution of all glaciers on Earth. *Nat. Geosci.* **12**, 168–173 (2019).

48. Wilson, R. et al. Glacial lakes of the Central and Patagonian Andes. *Glob. Planet. Change* **162**, 275–291 (2018).

49. Field, H. R., Armstrong, W. H. & Huss, M. Gulf of Alaska ice-marginal lake area change over the Landsat record and potential physical controls. *Cryosphere* **15**, 3255–3278 (2021).

50. Veh, G., Korup, O., Roessner, S. & Walz, A. Detecting Himalayan glacial lake outburst floods from Landsat time series. *Remote Sens. Environ.* **207**, 84–97 (2018).

51. Anderson, S. P. et al. Integrated hydrologic and hydrochemical observations of Hidden Creek Lake jökulhlaups, Kennicott Glacier, Alaska. *J. Geophys. Res.* **108**, 6003 (2003).

52. Milner, A. M. et al. River ecosystem resilience to extreme flood events. *Ecol. Evol.* **8**, 8354–8363 (2018).

53. Huss, M. et al. Toward mountains without permanent snow and ice. *Earths Future* **5**, 418–435 (2017).

54. Khanal, S. et al. Variable 21st century climate change response for rivers in High Mountain Asia at seasonal to decadal time scales. *Water Res.* **57**, e2020WR029266 (2021).

55. Nagy, T. & Andreassen, L. *Glacier Lake Mapping with Sentinel-2 Imagery in Norway* NVE Rapport 40-2019 (NVE, 2019).

56. Buckel, J., Otto, J. C., Prasicek, G. & Keuschnig, M. Glacial lakes in Austria—distribution and formation since the Little Ice Age. *Glob. Planet. Change* **164**, 39–51 (2018).

57. Chen, F. et al. Annual 30 m dataset for glacial lakes in High Mountain Asia from 2008 to 2017. *Earth Syst. Sci. Data* **13**, 741–766 (2021).

58. Mölg, N. et al. Inventory and evolution of glacial lakes since the Little Ice Age: lessons from the case of Switzerland. *Earth Surf. Process. Landf.* **46**, 2551–2564 (2021).

59. Iturrizaga, L. in *Encyclopedia of Snow, Ice and Glaciers* (eds Singh, V. P. et al.) 381–399 (Springer, 2011); https://doi.org/10.1007/978-90-481-2642-2_196.

60. Rickman, R. L. & Rosenkrans, D. S. *Hydrologic Conditions and Hazards in the Kennicott River Basin, Wrangell-St. Elias National Park Preserve, Alaska* Water-Resources Investigations Report 96-4296 (USGS, 1997); https://doi.org/10.3133/wri964296.

61. Bhambri, R. et al. Ice-dams, outburst floods, and movement heterogeneity of glaciers, Karakoram. *Glob. Planet. Change* **180**, 100–116 (2019).

62. Stone, K. H. Alaskan ice-dammed lakes. *Ann. Assoc. Am. Geogr.* **53**, 332–349 (1963).

63. Pfeffer, W. T. et al. The Randolph Glacier Inventory: a globally complete inventory of glaciers. *J. Glaciol.* **60**, 537–552 (2014).

64. Takaku, J., Tadono, T., Doutsu, M., Ohgushi, F. & Kai, H. Updates of 'AD3D30' ALOS global digital surface model with other open access datasets. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLIII-B4-2020**, 183–189 (2020).

65. Wulder, M. A. et al. The global Landsat archive: status, consolidation, and direction. *Remote Sens. Environ.* **185**, 271–283 (2016).

66. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).

67. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Chapman and Hall/CRC, 2020).

68. Gudmundsson, L. et al. Globally observed trends in mean and extreme river flow attributed to climate change. *Science* **371**, 1159–1162 (2021).

69. Blöschl, G. et al. Changing climate both increases and decreases European river floods. *Nature* **573**, 108–111 (2019).

70. Rokaya, P., Budhathoki, S. & Lindenschmidt, K.-E. Trends in the timing and magnitude of ice-jam floods in Canada. *Sci. Rep.* **8**, 5834 (2018).

71. Seneviratne, S. I. et al. in *IPCC Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) 1513–1766 (Cambridge Univ. Press, 2021).

72. Ng, F. & Liu, S. Temporal dynamics of a jökulhlaup system. *J. Glaciol.* **55**, 651–665 (2009).

73. Stan Development Team *Stan Modeling Language Users Guide And Reference Manual*, version 2.26.1; https://mc-stan.org (2022).

74. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).

75. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing* https://www.R-project.org/ (2022).

76. Clarke, G. K. C. Hydraulics of subglacial outburst floods: new insights from the Spring–Hutter formulation. *J. Glaciol.* **49**, 299–313 (2003).

77. Porter, C. et al. ArcticDEM. https://doi.org/10.7910/DVN/OHHUKH (Harvard Dataverse, 2018).

78. McNabb, R., Nuth, C., Kääb, A. & Girod, L. Sensitivity of glacier volume change estimation to DEM void interpolation. *Cryosphere* **13**, 895–910 (2019).

79. Hugonnet, R. et al. Uncertainty analysis of digital elevation models by spatial inference from stable terrain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15**, 6456–6472 (2022).

80. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**, 109 (2020).

Acknowledgements

We thank Planet Labs Inc. for a basic license in the Education and Research programme that granted free access to high-resolution Planet and RapidEye images. We thank many experts for sharing regional data on ice-dammed lakes: B. Einarsson, J. Carrivick and F. Tweed (Iceland); M. Huss and M. Werder (European Alps); M. Jackson (Scandinavia); P. Iribarren Anacona (Andes); R. Johnston (northwest North America); J. Steiner and V. Round (High Mountain Asia); and V. Foord and J. Jacklin for internal reviews.

Funding

Open access funding provided by Universität Potsdam.

Author information

Authors and Affiliations

1. Institute of Environmental Science and Geography, University of Potsdam, Potsdam-Golm, Germany

   Georg Veh, Natalie Lützow, Jenny Tamm, Lisa V. Luna & Oliver Korup

2. Institute of Geosciences, University of Potsdam, Potsdam-Golm, Germany

   Lisa V. Luna & Oliver Korup

3. Potsdam Institute for Climate Impact Research, Potsdam-Telegrafenberg, Germany

   Lisa V. Luna

4. LEGOS, Université de Toulouse, CNES, CNRS, IRD, UPS, Toulouse, France

   Romain Hugonnet

5. Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zürich, Zürich, Switzerland

   Romain Hugonnet

6. Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

   Romain Hugonnet

7. Federal Institute for Materials Research and Testing (BAM), Berlin, Germany

   Kristin Vogel

8. Ministry of Forests, Prince George, British Columbia, Canada

   Marten Geertsema

9. Department of Earth Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

   John J. Clague

Contributions

G.V. designed and led the study, analysed the data and produced all figures. N.L. and J.T. helped collect the data. R.H. and O.K. contributed to the data analysis. K.V., O.K. and L.V.L. contributed to formulating the Bayesian models. J.J.C., O.K., M.G., R.H. and L.V.L. contributed to the discussion of the results and to the writing of the manuscript.

Corresponding author

Correspondence to Georg Veh.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Guoxiong Zheng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 GLOF reporting between 1900 and 2021.

a, Histograms of all reported GLOFs (top), reported GLOFs with documented peak discharge Q_p (middle), and reported GLOFs with documented flood volume V_0. Numbers are the total number of GLOFs every three decades. **b**, Plot of year of GLOF occurrence versus year of first available document. White circles are individual GLOFs; colours are aggregated cases in 5-year intervals. **c**, Histogram showing the year of first available documentation for all cases in our database.

Extended Data Fig. 2 Regional trends of GLOF peak discharge Q_p with time *t* obtained by hierarchical quantile regression.

a, Temporal trends of Q_p for the periods 1900–2021 and 1990–2021. Colours distinguish between the median (blue) and 90th (orange) percentile of reported Q_p, including the posterior 95% HDI (shades). Numbers (n) and

percentages in lower left of panels are, respectively, the sample sizes and shares of GLOFs used in the quantile regression model for each region. **b**, Posterior trends in Q_p for the 50th and 90th percentiles. Posterior 95% HDI (lines) and median (dot) regression trend of Q_p are grouped by the periods 1900–2021 (dark green) and 1990–2021 (light green). Distributions overlapping with zero (thick black line) indicate the lack of a credible trend in Q_p. Percentages show posterior mass smaller (left) and larger (right) than zero.

Extended Data Fig. 3 Regional trends of GLOF volume V_0 with time t obtained by hierarchical quantile regression.

a, Temporal trends of V_0 for the periods 1900–2021 and 1990–2021. Colours distinguish between the median (blue) and 90th (orange) percentile of reported V_0, including the posterior 95% HDI (shades). Numbers (n) and percentages in lower left of panels are, respectively, the sample sizes and shares of GLOFs used in the quantile regression model for each region. **b**, Posterior trends in V_0 for the 50th and 90th percentiles. Posterior 95% HDI (lines) and median (dot) regression trend of V_0, grouped by the period 1900–2021 (dark green) and 1990–2021 (light green). Distributions overlapping with zero (thick black line) indicate the lack of a credible trend in V_0. Percentages show posterior mass smaller (left) and larger (right) than zero.

Extended Data Fig. 4 Global trends of GLOF volume V_0 and peak discharge Q_p with time t obtained by hierarchical quantile regression.

a, Pooled global trends in median and extreme V_0 for the period 1900–2021. **b**, Pooled global trends in median and extreme V_0 for the period 1990–2021. **c**, Posterior pooled trends in V_0 for the 50th and 90th percentiles for both periods. **d**, Pooled global trends in median and extreme Q_p for the period 1900–2021. **e**, Pooled global trends in median and extreme Q_p for the period 1990–2021. **f**, Posterior pooled trends in Q_p for the 50th and 90th percentiles for both periods. In **a**, **b**, **d**, and **e**, colours distinguish between the median

(blue) and 90th (orange) percentile of reported V_0 and Q_p, including the posterior 95% HDI (shades). In **c** and **f**, lines are the posterior 95% HDI and dots are the median of the global mean regression trends, grouped by the period 1900–2021 (dark green) and 1990–2021 (light green). Distributions overlapping with zero (dashed black line) indicate the lack of a credible trend in V_0 and Q_p. In **c** and **f**, π_{post} <0 denotes a posterior mass smaller than zero.

Extended Data Fig. 5 Posterior trends in local median (50th percentile) peak discharge Q_p and flood volume V_0 obtained by hierarchical quantile regression.

Posterior 95% HDI (lines) and median (dot) regression trend of **a**, V_0 and **b**, Q_p of ice-dammed lakes. Glacier names are those of the parent glacier (RGI ID). Lakes labelled ‘NA’ have no name. Distributions overlapping with zero (dashed black line) indicate lack of credible trend in V_0 and Q_p. Percentages are posterior masses smaller (left) and larger (right) than zero. Trends refer to $\log_{10}$-transformed values of V_0 and Q_p. The posterior densities show decadal trends in orders of magnitudes.

Extended Data Fig. 6 Regional seasonality of the monthly number of GLOFs and mean monthly air temperatures.

Mean monthly air temperatures at the locations of all lakes that caused a GLOF during the period 1901–2021 (data from CRU TS V 4.05, ref. [80]) were used to calculate the regional monthly average during that period. Months are sorted from 1 (January) to 12 (December).

Extended Data Fig. 7 Posterior differences in GLOF timing (doy) for 40 ice-dammed lakes obtained by Bayesian regression.

Lines and dots show, respectively, the posterior 95% HDI and median of the change in GLOF timing, expressed in days, between the first and the last reported GLOF from a given lake. Distributions overlapping with zero (thick

vertical line) suggest no credible change in the day of year (doy). Lakes labelled ‘NA’ have no name.

Extended Data Fig. 8 Regional elevation distribution of ice-dammed lakes.

Histograms in top row show all lakes (>0.01 km²) that currently exist in the six regions. Those in the bottom row are all lakes that burst out between 1900 and 2021. Vertical orange lines are the median elevation of ice-dammed lakes in each region.

Extended Data Fig. 9 Mean elevation change of glacier dams with reported GLOF magnitudes between 2000 and 2019.

Black line and blue shading are, respectively, the mean and standard error of glacier elevation change obtained by Gaussian Process regression[8]. Black ticks are dated GLOFs. Note that Tulsequah Glacier (RGI60-01.01521) dams two lakes. The reversed trend for Kyagar Glacier (RGI60-14.07098) is due to a glacier surge.

Supplementary information

Supplementary Tables

This file contains Supplementary Tables 1–6.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons

licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Veh, G., Lützow, N., Tamm, J. *et al.* Less extreme and earlier outbursts of ice-dammed lakes since 1900. *Nature* **614**, 701–707 (2023). https://doi.org/10.1038/s41586-022-05642-9

- Received: 03 March 2022
- Accepted: 07 December 2022
- Published: 15 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05642-9

Share this article

Anyone you share the following link with will be able to read this content:

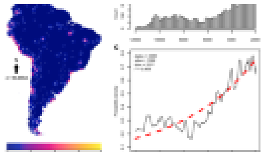
Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

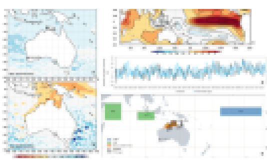
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05642-9



Widespread population decline in South America correlates with mid-Holocene climate change

09 May 2019

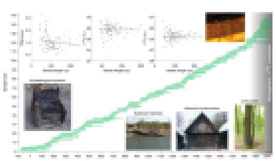
Philip Riris & Manuel Arroyo-Kalin



Compound climate extremes driving recent sub-continental tree mortality in northern Australia have no precedent in recent centuries

15 September 2021

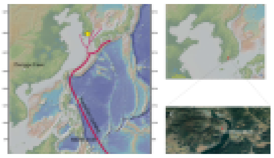
Kathryn J. Allen, Danielle C. Verdon-Kidd, … Patrick J. Baker



Recent European drought extremes beyond Common Era background variability

15 March 2021

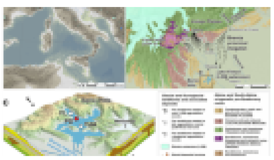
Ulf Büntgen, Otmar Urban, … Miroslav Trnka



[Abrupt Holocene climate shifts in coastal East Asia, including the 8.2 ka, 4.2 ka, and 2.8 ka BP events, and societal responses on the Korean peninsula]

25 July 2019

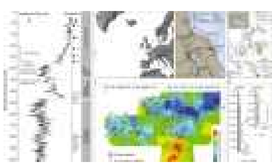
Jungjae Park, Jinheum Park, … Jieun Choi



[The fast-acting “pulse” of Heinrich Stadial 3 in a mid-latitude boreal ecosystem]

22 October 2020

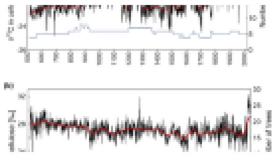
Federica Badino, Roberta Pini, … Stefano Benazzi



[The resilience of postglacial hunter-gatherers to abrupt climate change]

26 March 2018

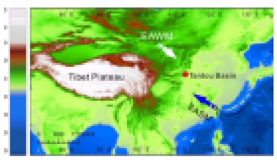
Simon Blockley, Ian Candy, … Nicky Milner



[Modern aridity in the Altai-Sayan mountain range derived from multiple millennial proxies](#)

11 May 2022

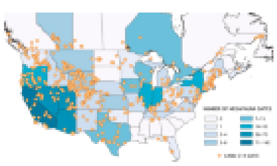
Olga V. Churakova-Sidorova, Vladimir S. Myglan, … Matthias Saurer



[New evidence of the emergence of the East Asian monsoon in the early Palaeogene](#)

28 November 2022

Li-Fei Su, Qian-Qian Zhang, … Cheng-Sen Li



[Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America](#)

16 February 2021

Mathew Stewart, W. Christopher Carleton & Huw S. Groucutt

- Article
- Open Access
- Published: 08 February 2023

Severe multi-year drought coincident with Hittite collapse around 1198–1196 bc

- Sturt W. Manning ,
- Cindy Kocik,
- Brita Lorentzen &
- …
- Jed P. Sparks

Show authors

Nature **volume 614**, pages 719–724 (2023)

- 21k Accesses
- 1 Citations
- 962 Altmetric
- Metrics details

Abstract

The potential of climate change to substantially alter human history is a pressing concern, but the specific effects of different types of climate change remain unknown. This question can be addressed using palaeoclimatic and archaeological data. For instance, a 300-year, low-frequency shift to drier, cooler climate conditions around 1200 bc is frequently associated with the collapse of several ancient civilizations in the Eastern Mediterranean and Near East[1,2,3,4]. However, the precise details of synchronized climate and human-history-scale associations are lacking. The archaeological–historical record contains multiple instances of human societies successfully adapting to low-frequency climate change[5,6,7]. It is likely that consecutive multi-year

occurrences of rare, unexpected extreme climatic events may push a population beyond adaptation and centuries-old resilience practices[5,7,8,9,10]. Here we examine the collapse of the Hittite Empire around 1200 bc. The Hittites were one of the great powers in the ancient world across five centuries[11,12,13,14], with an empire centred in a semi-arid region in Anatolia with political and socioeconomic interconnections throughout the ancient Near East and Eastern Mediterranean, which for a long time proved resilient despite facing regular and intersecting sociopolitical, economic and environmental challenges. Examination of ring width and stable isotope records obtained from contemporary juniper trees in central Anatolia provides a high-resolution dryness record. This analysis identifies an unusually severe continuous dry period from around 1198 to 1196 (±3) bc, potentially indicating a tipping point, and signals the type of episode that can overwhelm contemporary risk-buffering practices.

Main

The vast Hittite Kingdom and subsequently Empire, based in central Anatolia, Turkey, with its capital at Hattusa, is recognized from both rich archaeological remains and textual sources as one of the major Old World powers of the Eastern Mediterranean and Near East between 1650 and 1200 bc. At its apex, the Hittite Empire maintained control over central, southern and southeastern Anatolia, the northern Levant and northern Syria, with almost all of Anatolia being under the Hittite sphere of influence (Extended Data Fig. 1). During this time, the Hittite Empire vied with the Egyptian Empire for sociopolitical dominance in the Near East, a struggle that culminated in the largest battle of the era at Kadesh in Syria in the early 13th century bc[15].

Around or shortly after 1200 bc, the Hittite Empire and central administrative system collapsed in a great realignment that reverberated around the Near East[4,11,12,13,16,17,18,19,20]. The reign of the last known king, Suppiluliuma II, began around 1207 bc and included claimed victories against several intra-Anatolian rivals (Wiyanawanda, Masa, Lukka and Ikkuna) and Alashiya (Cyprus) in sea and land battles, but no further Hittite rulers were recorded subsequently. An inscription of the Egyptian ruler Ramesses III—approximately dated to 1188 bc or 1177 bc, depending on

selection and debate in Egyptian history and chronology—lists the Hittites among those swept away by the ‘Sea Peoples’ before they attacked Egypt[4,11,17,18].

The end of settlement at Hattusa itself has been a key topic of historical scrutiny. Long considered a victim of attack, whether by the Sea Peoples or local Anatolian raiders, archaeological investigations now indicate that the city was abandoned and emptied by the royal administration and only later burnt[11,12,16,21,22,23]. Hattusa was the centripetal political and core religious venue of the Hittite gods and kings for centuries, and the reasons for its abandonment remain unclear. The Hittite central sociopolitical and economic system withstood multiple, diverse crises during its tenure: intra-Anatolian tensions with Kaska invaders, rivalries and tensions among constituent elements of the Empire (Arzawa-Mira, Tarhuntassa and Karkemish), dynastic politics (rivals and usurpers to the Great King), and threats such as plague, as well as external challenges, are regularly evident in the Empire’s history[11,12,13,14,15,16,21,22,23,24,25]. However, the final collapse and abandonment of Hattusa and the central Hittite administration (and thus the cessation of written historical documents) was different and seismic in its scale of impact. Despite evidence of adaptation and continuation of elements of Hittite political and cultural systems at various loci within former Hittite territories (especially in the southeast and in northern Syria)[26,27,28], it is undeniable that the succeeding neo-Hittite states that emerged were of a very different, much smaller scale and character[19,24,25,26,27,28,29]. As Middleton recently stated regarding Hattusa’s collapse: “A context of conflict was nothing new … and so it seems appropriate to conclude that something historically specific (and perhaps never recoverable) happened to precipitate the abandonment”[22].

Recent scholarship inspired by contemporary concerns around climate change has increasingly shifted from explaining the fall of the Hittites and the wider collapse of several Late Bronze Age civilizations with invaders or raiders, earthquakes or various political–economic changes as prime movers, to speculate instead on a possible underlying climatic or environmental driver[1,2,3,4,11,30,31]. Such work cites a variety of palaeoclimatic proxy evidence indicating a probably drier and cooler regime in the Eastern Mediterranean–Anatolian–Near East region in the period from the 13th to

10th centuries bc. However, this evidence is often only loosely placed in absolute temporal terms[20] and provides a record of (at best) low-frequency climatic change. In the Hittite case, drought, famine, reduced harvests and concomitant reductions in workforce and military strength might be suggested as possible explanatory structures[32], mirroring drought as the main threat to agriculture and food security in Anatolia in more recent history[33,34,35,36]. Several texts from the 13th century bc refer to apparent grain shortages or famine in Hittite lands. However, the interpretation of this material lacks detail and context[20,37], whereas other critical assessments of the data from 13th century bc Hattusa do not necessarily indicate pending crisis[38]. Nonetheless, it may be observed that the Hittite Kingdom and especially its main centre, Hattusa, engaged in major landscape clearance[39] (resulting in soil erosion) and specialized pastoralism[40], and relied on higher-risk subsistence strategies dependent on water. Cereal agriculture both locally and from its periphery supplied the necessary large volumes of grain—attested in the early Hittite period by the discovery of a vast storage silo and subsequently via additional, smaller, dispersed silos[41]. In combination, these linked strategies may have increased production but amplified risk. As in several other early agrarian states, cereals formed a key subsistence and tax base[42]; in reverse, such states would have been vulnerable to serious and sustained threats to their cereal (and other crop) harvests and animal husbandry.

However, absent from such Late Bronze Age environmental resilience and sustainability assessments are nearly absolutely dated and highly resolved (that is, annual-scale) climate indicators for this region generally, and specifically from the Hittite administrative core in central Anatolia, capable of defining the nature of any critical climatic forcing potentially relevant to the collapse of the core Hittite administrative centre at multiple temporal scales. Gradual, low-frequency, shifts in climate whose amplitude does not completely alter the area's bioclimatic system—that is, the shift attested in existing palaeoclimate archives for the period[1,2,3,4,30,31,43]—are less likely to undermine human strategies based around adaptation and resilience (through a wide variety of strategies of diversification, storage and social networking[5,6,7,8,44]). Similarly, one-year droughts (or comparable high-frequency challenges) are expected, particularly in a semi-arid region such as central Anatolia, where periodic and even regular droughts are anticipated

as the main threat to agriculture[8,10,33,34,35,36,44]. Traditional farming practice (and storage strategies) in the greater Mediterranean region, and most agrarian cases, is adapted with the aim to be able to cope with one bad year[5,8,9,10,34,44]. What breaks the system—turning regular instances of food scarcity into famine and crisis—is multiple, consecutive harvest failures over two and especially more years[5,9,10,44]. On the basis of various records from the past several hundred years, such multi-year consecutive serious droughts leading to prolonged harvest reduction and failure (and so famine and associated threats from disease to violence) are rare but probably of potential historical relevance in central Anatolia[10,33,34,35,36,37,45,46]. Recognition of such critical episodes requires annually resolved data that facilitate multi-scalar (including high-frequency) climatic assessments. Here, high-resolution tree-ring records from central Anatolia enable us to examine climate change in the Late Bronze Age in decadal to multi-decadal generalities, but also with annual-scale and historical specificity. We use data derived from nearly absolutely placed tree-ring time series—both ring-width based and from ^{13}C stable isotope analysis—to characterize an annually resolved climate record of moisture availability for the period around 1500–800 bc and investigate whether there was a catastrophic continuous multi-year drought episode of historical relevance across central Anatolia around 1200 bc.

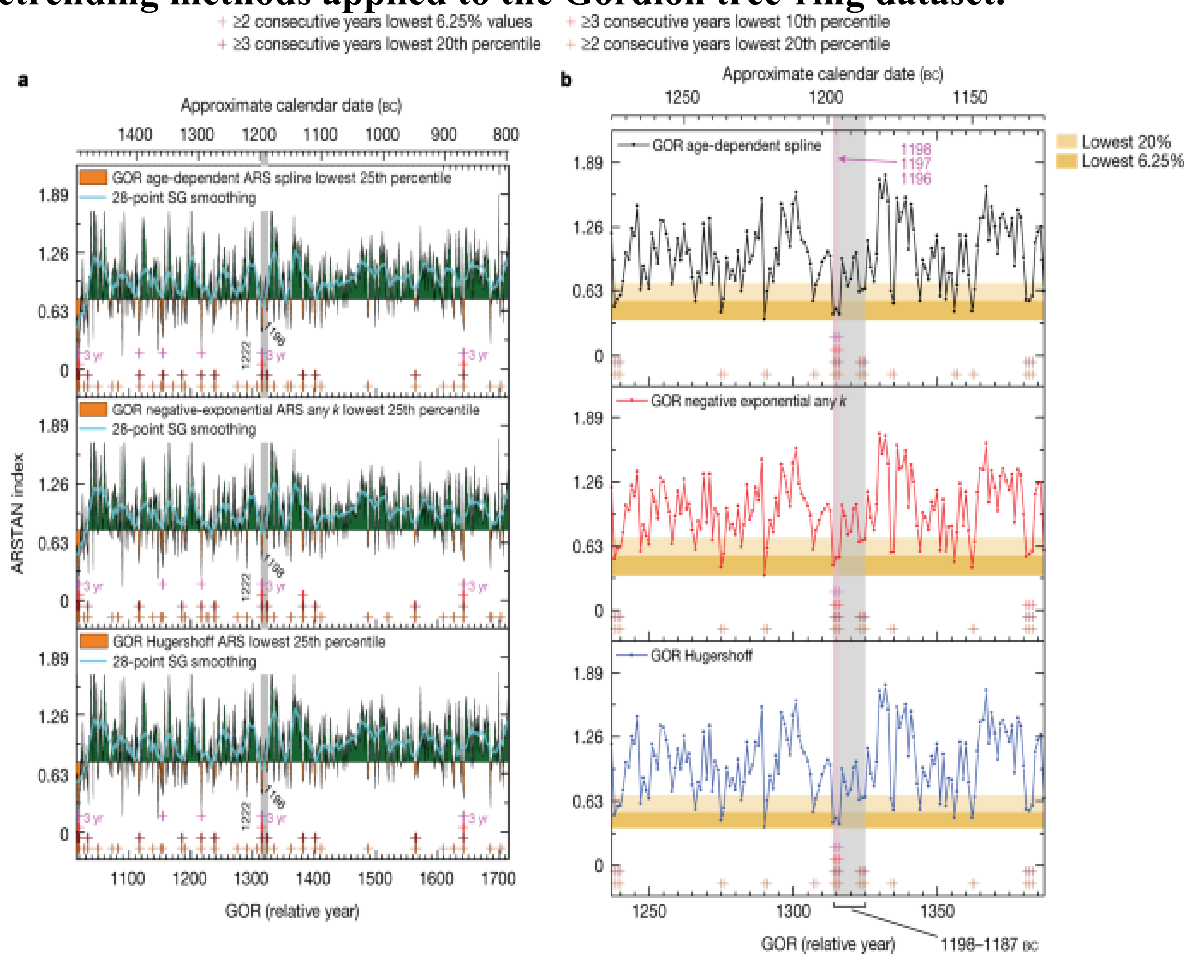
Juniper (*Juniperus excelsa* and *Juniperus foetidissima*) timbers recovered from archaeological excavations at the site of Gordion in central Anatolia, about 230 km west of the Hittite capital Hattusa near the western frontier of the land of Hatti (the core kingdom of the Hittites), provide an annually resolved tree-ring chronology from around 1775–748 (±3) bc that traverses the end of the Late Bronze Age (Extended Data Fig. 2 and Methods). Twenty-three samples, representing around 18 different trees including the year 1200 bc, form a robust tree-ring chronology. Investigations of recent juniper populations in Anatolia and central Asia indicate that trees, especially from lower elevations (like those likely to be used at ancient Gordion), are susceptible to and exhibit reduced growth owing to decreases in late spring–summer precipitation[33,46,47,48] (also critical to cereal production in Anatolia). Thus, examination of the tree-ring patterns in the Gordion tree-ring chronology and identification of episodes of substantially reduced growth increments (narrower tree rings)—after removing effects of

age-related growth trends and stand dynamics from the series, and in the absence of any indications of fire or insect attack—probably indicates drier years, with extremes marking probable drought episodes critical to agricultural production and subsistence.

We investigated the Gordion chronology, detrending the individual tree-ring width series to derive a maximized climate signal, which we compared with modern analogous climate data for the central Anatolian region (Extended Data Figs. 1, 3–7 and Methods). Taking the lowest 25% of growth values in the Gordion chronology as indicating drier years, these events occur regularly between around 1497 and 797 bc (Fig. 1 and Extended Data Figs. 5 and 8a). For comparison, the Polatlı meteorological station, which offers the nearest modern climate analogue to ancient Gordion (17 km away), records less than 300 mm of annual precipitation for approximately 20% of the period from ad 1929 to ad 2009. Total annual precipitation of 300 mm is regarded as an approximate minimum threshold for a viable wheat harvest[44]. Only 6.25% of years from ad 1929 to ad 2009 recorded less than 250 mm of annual precipitation, less than the amount generally regarded as the minimum for cereal cultivation in the Near East, and an amount that would probably produce serious harvest reduction or failure[44,49]. Although regional precipitation values vary, the largely semi-arid central Anatolian region demonstrates overall congruent records or trends (Methods). In particular, recent and historical instances of extreme arid years are usually common to much of the greater central Anatolian or core Hittite region[33,35,36] (Extended Data Figs. 6 and 7). Thus, if we regard the relative occurrence ratio of drier years in the recent Polatlı record as indicative for the broadly similar period of the drier later second millennium bc in central Anatolia[30], and use the lowest 20% of values as indicating probable substantial reductions in the harvest, then the Gordion chronology contains only 80 or 85 out of 701 years (11.4%–12.1%) that are part of two or more consecutive such dry years. Between 1270 and 1135 bc, when we also have our best sample replication with 10–18 trees per year, there are only six such sets of years. Only 13 or 16 years (1.9%–2.3%) in total are part of two or more consecutive drought years with the lowest 6.25% of values, representing probable serious harvest reduction or failure. Between around 1270 and 1135 bc there is only one interval in these 135 years—1198–1196 bc—with two or more consecutive years in the lowest 6.25% of values (three

consecutive years). Further, in the 12-year period from 1198 to 1187 bc, there were between 6 and 8 years (50–67%) in the lowest 20% of values (Fig. 1 and Extended Data Fig. 8a). Smoothing the data with a 28-point Savitzky–Golay filter representing an average human generational timeframe[50], the period around this time represents either the driest or second-driest multi-year interval between 1400 and 1000 bc. This extremely dry interval stands out as a probably substantial climate challenge to food production and subsistence in central Anatolia that may have defeated normal strategies and storage provision in the Hittite administrative core. The dates—approximately 1198–1196 bc—are compatible with the historically derived timeframe of Hittite collapse and reorientation[4,11,12,13,16,17,18,19,20,21,22,23,32], and lend an historical specificity that is usually lacking in general low-frequency arguments suggesting linkages between climate and history.

Fig. 1: Proxies of drier to drought climate from three different detrending methods applied to the Gordion tree-ring dataset.

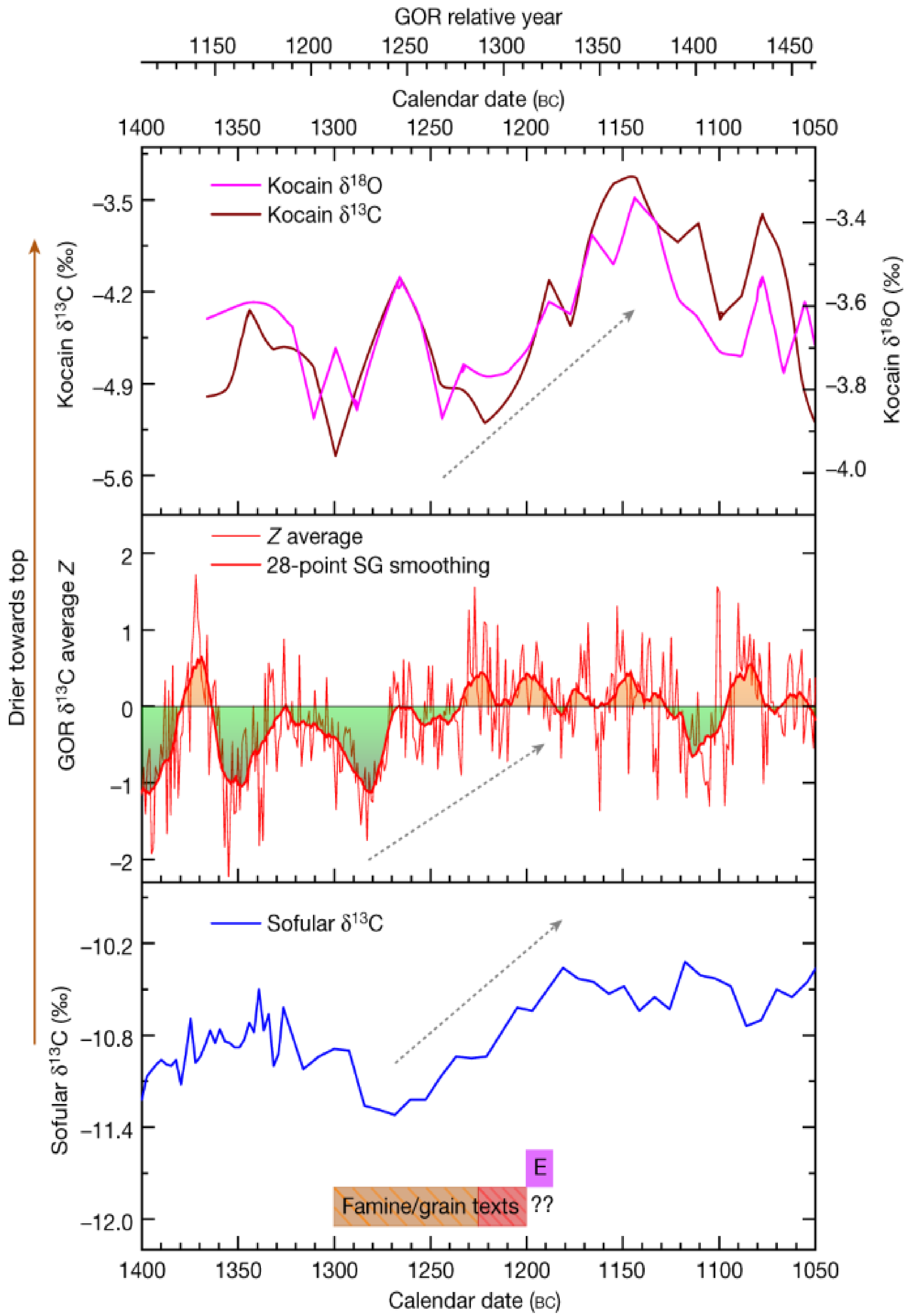


a, Tree-ring record from 1497–797 bc (Methods). The driest 25% of years is shown in orange (other 75% of years in green shading); division indicated by black horizontal line; 28-point Savitzky–Golnay (SG) filter is shown. Instances of 2 or 3 consecutive dry to very dry years at various levels are indicated. There are 3 instances of the driest 6.25% of years occurring consecutively (1494–1492 bc, 1198–1196 bc and 871–869 bc). The grey bar indicates the 12-year period from 1198 to 1187 bc with 3 consecutive years from 1198–1196 bc in the lowest 6.25% of all years and with 6 or 7 years (50–58%) in the lowest 20% of values. GOR, Gordion. **b**, Close-up of the period 1275–1125 bc, showing annual ARSTAN (ARS) index values (Methods) highlighting those in the lowest 20% and the lowest 6.25% of values.

An alternative approach to recovering proxy moisture availability levels from tree rings at lower elevations in a semi-arid area such as central Anatolia is via ^{13}C stable isotope analysis, with higher $\delta^{13}C$ values usually indicating drier conditions. The $\delta^{13}C$ is defined as $[(R_{sample} - R_{standard})/R_{standard}] \times 10^3$, in which R_{sample} is the ratio of ^{12}C to ^{13}C in the unknown and $R_{standard}$ is that same ratio in the internationally agreed-upon standard PeeDee Belemnite (PDB). The $\delta^{13}C$ is driven by the gradient of CO_2 between the atmosphere and the interior of the tree leaf, which is controlled predominantly by leaf stomata, the leaf apertures that control water loss and CO_2 transport into the tree. Therefore, $\delta^{13}C$ values fluctuate in response to changing moisture availability. We compare average *z*-transformed results (where positive values indicate higher $\delta^{13}C$) from a series of ^{13}C measurements on α-cellulose extracted from tree rings of four of the Gordion trees with stable isotope records from the Sofular Cave (northwest Anatolia) and Kocain Cave (southwest Anatolia) for the period 1400–1050 bc (Fig. 2 and Methods). The data show a noticeable gradient towards drier conditions in the later 13th century bc, then probable consistently dry to very dry conditions as exhibited by the Gordion tree rings across the interval 1232–1192 bc with drier spikes from 1222 to 1221 bc and around 1195 bc, aligning closely with ring-width minima (Fig. 1 and Extended Data Fig. 8a). This period of drier gradient matches the timeframe of the ancient texts indicating grain shortages in Hittite lands[11,20,37] and

peaks around the time of the Hittite collapse, consistent with the tree-ring width analysis.

Fig. 2: $\delta^{13}C$ record from Gordion tree rings compared with $\delta^{13}C$ and $\delta^{18}O$ records from Sofular and Kocain Caves, Turkey, and the period referred to in texts mentioning famine or grain shortage in Hittite lands.


GOR relative year
1150
1200
1250
1300
1350
1400
1450
Calendar date (BC)
1400
1350
1300
1250
1200
1150
1100
1050
Kocain δ18O
Kocain δ13C
Kocain δ13C (‰)
−3.5
−4.2
−4.9
−5.6
Kocain δ18O (‰)
−3.4
−3.6
−3.8
−4.0
Z average
28-point SG smoothing
GOR δ13C average Z
2
1
0
−1
−2
Drier towards top
Sofular δ13C
Sofular δ13C (‰)
−10.2
−10.8
−11.4
−12.0
E
Famine/grain texts
??
1400
1350
1300
1250
1200
1150
1100
1050
Calendar date (BC)

$\delta^{13}C$ and $\delta^{18}O$ records from Sofular (bottom) and Kocain (top) Caves compared with the Gordion overall combined average *z* score $\delta^{13}C$ time-series, $\delta^{13}C_{CorZ}$, chronology (middle). The Gordion chronology is also represented smoothed with a 28-point Savitzky–Golnay filter (middle). *Z*-transformed values greater than zero (drier) are highlighted with orange shading above the line and contrasted with green shading for wetter conditions below the line. Dryness increases towards the top of the graph. Bottom, E indicates the End Hattusha–Hittite Kingdom period during the reign of Suppiluliuma II and before Ramesses III year 8. Bottom, out of 9 contemporary texts that refer to famine or grain shortage in Hittite lands[20,37], 4 are from the 13th century bc, 3 are undated but also probably originate from the 13th century bc (orange shading), and 2 are from the end of the 13th century bc (red shaded area). The dashed arrows highlight indications of a drying trend in each record.

Discussion

Critiques of many attempts to associate climate change with specific historical or archaeological shifts have focused on the often simplistic and reductionist logic, noting that there is frequently little consideration of how the climate shift is likely to have affected the relevant human subsistence and wider social, economic and political systems[6,7,51]. In several cases, the timeframes on one side or the other are not sufficiently highly resolved to enable a discussion of whether a specific correlation might exist—let alone whether such a specific correlation might in fact provide an element of causation[20,52]. Climate alone does not create or cause history. Rather, it is one of the forces comprising the context (*habitus*) in which human and other actors and vectors make decisions, interact, live lives and affect their surroundings. Humans, especially those living in semi-arid circumstances, expect both variable weather and year-to-year climate differences, and develop practical and physical adaptations (such as agricultural strategies, storage and water infrastructure) and social life (forms of networking) to create resilience against realistic and anticipated scenarios[5,8,34,44].

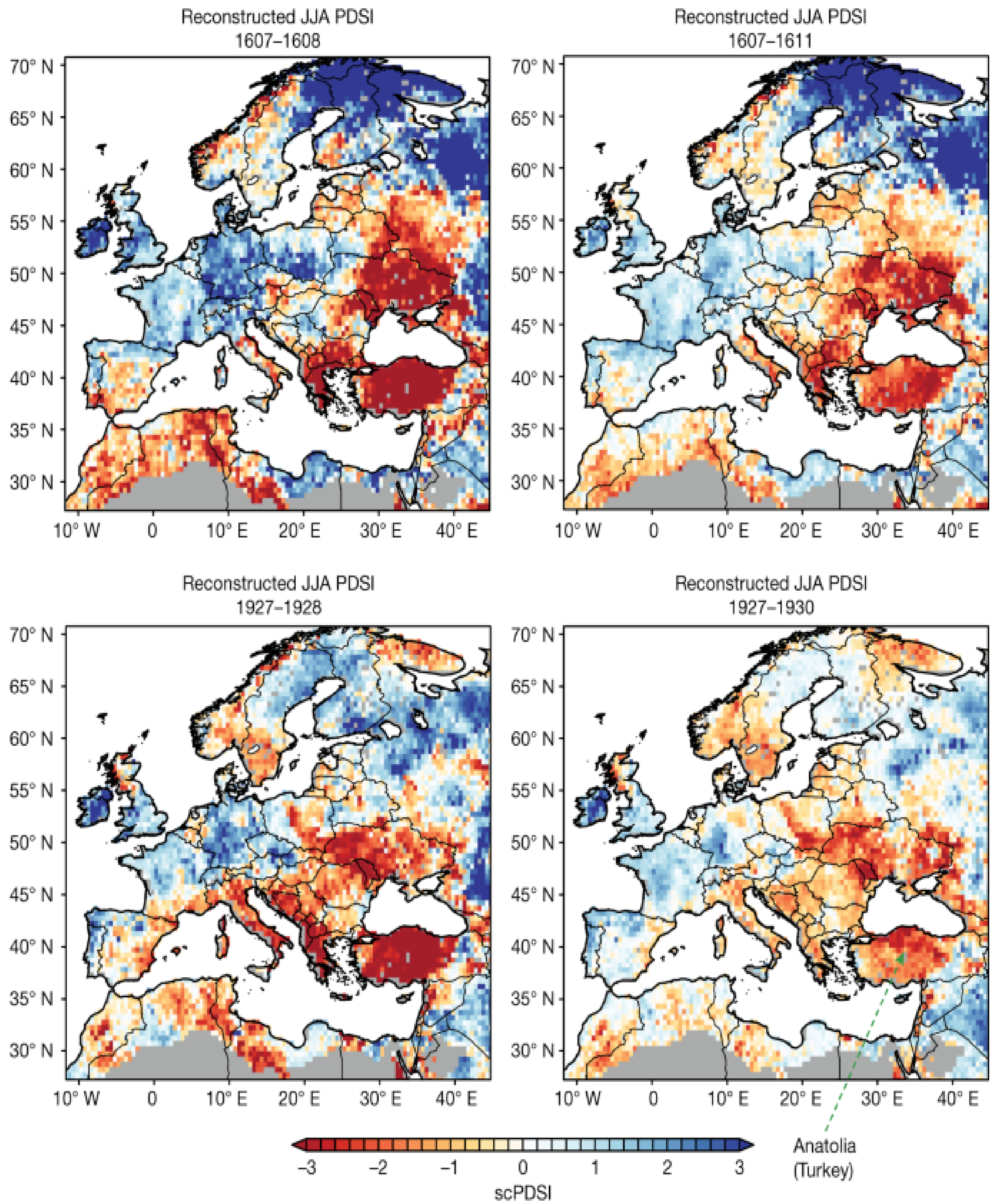
As noted, the Hittite Empire and its core were accustomed to drought threat or stress. Evidence of military installations and political iconography point

to efforts to manage territory and landscape[11,12]. Major water management infrastructure (such as dams and reservoirs) and their strong ritual associations point to efforts to control water resources and ameliorate deficits[53,54]. However, as observed for the medieval Middle East, “even well-organized regimes found it hard to cope with long periods (more than two years) of food shortages”[55]. Such periods were not anticipated. In Late Bronze Age Anatolia, although dry years and droughts were a regular feature, and even pairs of arid years generally (although not always) occurred at least once a generation, major droughts causing serious harvest failure in consecutive years were much rarer, probably occurring only once or at most twice in a century (Fig. 1 and Extended Data Fig. 8a). This is similar to observations from the second millennium ad[10,33,35,36,45,46]. In this context, the evidence of a serious drought and major harvest reduction in central Anatolia for three consecutive years between 1198 and 1196 bc, towards the end of a period from 1204 to 1192 bc according to tree-ring width data (Fig. 1) or from 1230 to 1192 bc according to the $\delta^{13}C$ data (Fig. 2) indicating generally lower moisture, is probably an historically important episode that would have severely challenged existing adaptations and resilience strategies. Large land-locked centres and territories reliant on regional grain production and specialized animal husbandry (such as Hattusa and the core Hittite realm, along with the Empire’s overall subsistence and tax base infrastructure) may have been particularly vulnerable in such circumstances (despite various other loci being able to adapt and survive). Outside a few peripheral maritime or riparian locations, marshalling substantial long-distance shipments of bulk subsistence products—if available—would have been logistically impractical for central Anatolia given the over-land transport technology of this period (over-land pre-modern caravan routes of the region centred on items of low bulk and higher value per unit volume or weight[56]). Much of the Hittite heartland would have been effectively isolated and forced to survive on local resources, and, consequently, in crisis as these were progressively exhausted across one, then two, and finally three years of consecutive serious drought. Further, textual evidence hints that surrounding entities, and especially some of the maritime linked centres (such as Ugarit)—which had their own difficulties[20,37]—withheld possible grain shipments, in turn exacerbating crises in Anatolia. Intra-Anatolian conflicts may have been stimulated by struggles for key subsistence resources in and immediately following this

period. Such circumstances would further stress underlying political, economic and social fault lines within the Hittite world that had apparently been building during the 13th century bc[21,23], and around it, and also provide the context for disease outbreaks. Thus, we propose that these three years, between around 1198 and 1196 bc (and the period from 1198 to 1187 bc), may well mark and form a key part of the circumstances that precipitated the collapse of the Hittite Empire. However, we also must acknowledge that we lack the evidence to establish direct causation. It remains likely that this rare, disastrous climate episode from 1198 to 1196 bc tied in with (or enabled) other, probably human, forces that merit further examination. This drought event thus contributed to, but did not solely cause, the collapse and break-up of the Empire.

The Hittite collapse forms part of a wider set of changes occurring across the Old World around 1200 bc[1,2,3,4,17,18,19,20]. Climate alone was not the sole cause of these changes; very different histories are evident within the greater region[20,57] (Methods). Nonetheless, if we consider the two instances of major back-to-back drought widely attested in Anatolia in the past seven centuries[10,36,45,46], alongside data and analysis available from the tree-ring width proxy reconstructions in the Old World Drought Atlas[58], we find that such severe drought tends to affect most of the area of the Hittite Empire, albeit with some variation in drought intensity (Fig. 3 and Extended Data Figs. 6 and 7). At the same time, as evident from these cases, other areas in the greater Mediterranean–Near East may well have experienced differing circumstances. Such strong differentials, with decisive adversity affecting an adjacent zone, may have helped to fuel population movements, reorientations of trade and political fragmentation, leading to new political alignments and configurations (and not just blanket collapse), and thus also form part of the explanation for the major regional reorientation in the early 12th century bc in conjunction with severe drought in Anatolia[20,57,59].

Fig. 3: Reconstructed summer dryness levels across Europe and the Mediterranean region around the consecutive major drought years ad 1607–1608 and ad 1927–1928.



Climate proxy data from the Old World Drought Atlas[58] (http://drought.memphis.edu/OWDA/) derived from tree-ring measurements, showing self-calibrating summer (June–July–August (JJA)) Palmer Drought Severity Index (scPDSI) values (negative values indicate drier and positive values indicate wetter) across the Old World region during and around the

consecutive major drought years ad 1607–1608 and ad 1927–1928 (2 consecutive years)[10]. Our study suggests that there were 3 consecutive drought years from 1198 to 1196 bc in Anatolia.

Low-frequency climate change creates long-term general forcing conditions affecting human evolution and strategies of adaptation and resilience. However, it is the potentially critical vulnerability of established human systems to unexpected and consecutive multi-year extremes, with concomitant combinations of stresses, that can break and overwhelm established adaptations and resilience practices—and greatly amplify the effects of high-risk land management practices and degradative land use. This applies in history as well as the present in the face of current climate change. The probable multi-year major drought that we identify occurring from 1196 to 1198 bc in Anatolia offers a salient example.

Methods

Proxy climate data and history of climate and society

It is important to acknowledge that (1) all proxy climate reconstructions are both approximate and limited; and (2) the human dimensions of effects and responses to climate (and other) change are complicated with internal diversity and varying systems from bottom to top (of any hierarchy) within societies and networks of external connections affording plural possible trajectories[6,7]. Climate anomalies or changes by themselves do not cause human societies to rise or fall. They are part of the context within which individuals, groups, and larger social, economic, and political entities make decisions, act, and react[5,6,7,16,22,26,28,34,45,51,57]. As argued in the main text, multi-year (thus prolonged and extended) climate (or environmental) challenges[60], such as unusual dryness in a semi-arid context (that is, drought), create circumstances that especially challenge existing systems. There is no simple correlation of crisis equalling collapse; indeed, in history there are few cases of real total collapse, rather, there is transformation (adaptation)—for example, the end or marked change of an existing political system or hierarchical model (such as the end of an empire or state)—to accommodate changing circumstances and usually with considerable continuity in many of the other, underlying, aspects of life[6,7,16,40,59,61]. The scale or harshness of the change (the 'collapse') in such transformations depends on many factors, from the severity of any climate driver to the constitution and history of the relevant society. In particular, for circumstances in which a 'rigidity trap' applies, as a previously successful, long-term system becomes entrenched and homogenous across a large area, there may be a harsh transformation[62].

The Hittite case addressed in this Article engages with all these issues. We identify, but cannot exactly quantify, both (1) a period of increasing dryness (during the 13th century bc) followed by several decades of dry circumstances (late 13th to early 12th centuries bc—which is not necessarily the most extreme in the record but nevertheless a conspicuous trend) (Fig. 2 and Extended Data Figs. 9 and 10), and (2) within the latter

dry period, a trio of consecutive years with unusually low tree-growth indicating very dry conditions that probably created extreme stress on subsistence agriculture and food storage strategies (Fig. 1 and Extended Data Fig. 8a) (see 'Tree-ring data' and 'Stable isotope data' below), which seems coincident with and part of the circumstances in which the large political entity of the Hittite Empire (probably falling into a rigidity trap context by this time) suffered a harsh transformation. At the same time, this collapse was a dynamic transformation, and many other aspects of society, economy and politics continued in the region (see further below)[20,22,26,27,28,29,40,57,59].

Tree-ring data

A tree-ring chronology comprising *Juniperus* sp. samples (*J. excelsa* and *J. foetidissima*) from archaeological contexts at Gordion, central Anatolia, and which were probably procured from the surrounding region, has previously been constructed using standard methods[63,64,65] based on tree-ring width analysis and discussed in terms of locus, site context, and near-absolute dating[66,67,68,69]. These trees grew in the period before substantive human modification of the wider Gordion landscape (in the Iron Age to the Roman period[34,49,70,71]), probably on hillsides of lower to moderate elevations, and thus can be expected to reflect the area's natural environment and climate. While there is no modern juniper population remaining in the Gordion region to act as a direct analogue to the archaeological material, recent studies of the same juniper species growing in neighbouring regions produce positive growth responses to greater spring–summer moisture availability and negative responses to increasing temperature[33,46,47,48,72]. We do not attempt to quantify proxy precipitation or moisture availability given the lack of a substantial modern analogue in the immediate region around Gordion. Both available modern analogue juniper trees and long-term meteorological (pre-1950s) data are absent. Instead, we seek to characterize indications of unusually dry conditions, both from instances of unusually narrow (reduced) tree-ring growth, consistent with modern occurrences of dry and/or hot conditions[33,46,47,48,72], and via a shift in stable ^{13}C values only plausibly explained by reduced moisture availability (see 'Tree-ring data' and 'Stable isotope data' below). Thus, we offer what we

may term ‘moisture sensitive proxies’ and identify when these appear to indicate a multi-year unusually dry interval (see further below).

The Gordion juniper chronology runs from relative years 737–1764 or approximately 1775–748 ±3 bc (an alternative dating scheme would place the chronology approximately 3 years later or more recent[73]). Twenty-three *Juniperus* sp. samples, probably representing 18 trees include the year ~1200 bc (Extended Data Fig. 2). These were selected for the current project. Using the quality control COFECHA software (version 6.06P)[74], the samples crossdate to form a robust chronology (Extended Data Fig. 2). All samples crossdate well with long (>100 years) overlaps with at least several other samples within the series set. The minimum acceptable Baillie–Pilcher crossdate *t*-value (t_{BP})[75] employed in COFECHA is 3.5. For long (>100 years) crossdates this (somewhat arbitrary) value stands up well as an approximate threshold[76]. However, in reality, many years of dendrochronological work has shown that higher t_{BP} values are often necessary to produce secure crossdates[77,78]. Our Gordion series have a much higher mean t_{BP} value of 8.98 (and mean $t_{BP} = 9.17$, when four tree-ring series with instances of $t_{BP} < 3.5$ against the rest of the chronology are excluded), with the average of all crossdated overlaps covering 304 years (and 309 years removing the 4 cases where $t_{BP} < 3.5$). In the whole series only two overlaps (n) are <100 years (one at $n = 74$ years with $t_{BP} = 4.4$, and one $n = 71$ years with an unsatisfactory $t_{BP} = 3.3$). The series intercorrelation, measured as a Pearson correlation coefficient (r), is $r = 0.632$. Ring-width data employed are listed in Supplementary Table 1. COFECHA did not identify any errors in which all or part of a ring-width segment had higher correlation against the rest of the site chronology other than at the dated position. With the default COFECHA settings (segment length 50 years, lagged 25 years) COFECHA finds 29 outlier segments with low correlation but, if the segment length is increased as appropriate in a case like this with long series lengths and relatively complacent data[65,74], then the number of flagged outliers reduces substantially, consistent with a robust chronology.

In order to derive a likely climate signal (moisture sensitive proxy) we used the ARSTAN software (version 49v1b_MRWE)[64,79,80] to standardize the

chronologies (Supplementary Table 1) by eliminating age-related growth trends and other possible non-climatic influences on tree-ring growth such as stand dynamics and local disturbances. Given the characteristics of the raw ring measurement profiles (Extended Data Fig. 2), which include a general negative exponential growth trend in most cases, some instances of initial suppressed growth, and generally long, largely complacent mature growth phases, we considered six detrending options (we did this to ensure or show that the unusually low tree-ring growth in the years ~1198–1196 bc is a consistent and robust observation across a reasonable ensemble of analytical approaches[81]). These comprised the following ARSTAN standardization curves: (1) the Melvin age-dependent spline[82]; (2) negative exponential curve (any *k*); (3) Hugershoff growth curve (except one sample, GOR-3, where warning that ‘itmax exceeded in amoeba’, and hence we used the age-dependent spline for this sample); (4) a 50-year spline with 50 per cent variance cut-off[83]; (5) the Friedman variable super smoother[84] with mid $\alpha = 5$ setting; and (6) a relatively stiff spline with a 100-year cut-off, given the form of the data and the research aim to investigate higher versus low-frequency variations (observed growth versus expected growth) (Extended Data Figs. 3–5). We employ approaches 1–3 in Fig. 1. The same analysis using the other three detrending versions (Extended Data Figs. 3–5) is shown in Extended Data Fig. 8a. All analyses produced similar results.

The data were combined into an overall chronology with a bi-weight robust mean including use of the weighted Rbar stabilization method[85]. This reduces the influence of extreme values within the series[79]. ARSTAN then produces three chronology versions, a Standard (STNDRD) detrended version, a Residual (RESID) version which reuses the Standard chronology after autoregressive modelling of the detrended series, and an ARSTAN (ARS) chronology which combines the pooled autoregression (persistence common and synchronous among a large proportion of tree-ring series and therefore probably climate related) into the RESID version and is anticipated to express the strongest climate signal possible[79]. We employed the ARSTAN chronology for our analyses showing the residual or difference (rt − gt, where rt = observed growth and gt = expected growth) ARSTAN chronology (Fig. 1, Extended Data Figs. 4, 5, 8). The mean inter-series correlation coefficient (Rbar) and the expressed population signal (EPS) were determined using 50-year moving windows with 25-year

overlaps to assess the strength and reliability of the chronology in representing the original source population's mean growth pattern over time (we assume that the juniper trees are probably from a similar source area, as they were cut and employed for the same specific construction project[67], and exhibit high correlation, but this is an hypothesis). The EPS values for the analyses exceeded the arbitrary but commonly cited threshold value of 0.85 regarded as suitable for dendroclimatological purposes[86] between relative years 1015–1715 (~1497–797 bc) (Extended Data Fig. 5), and this is the period employed in our analyses (Fig. 1, Extended Data Fig. 8a). In our study, given the absence of available recent juniper data in the immediate region, it is not possible to quantify the relationship between relevant narrowest tree rings and then the detrended ARSTAN values per year (for juniper in the Gordion area) versus precipitation. Nonetheless, as a reasonable analogue, we might consider the relationships between juniper in Southwest Anatolia versus instrumental May–June precipitation (the critical months for a precipitation proxy in this region)[46]. Here, ring-width data are available for three juniper chronologies[87] (https://www.ncei.noaa.gov/access/paleo-search/study/6301, accessed August 2022). We can compare ARSTAN modelling of these using the Melvin age-dependent spline[82] (see Extended Data Fig. 8b). Observed meteorological data 1931–2000 find 10 years (14.3%) with ≤30 mm May–June precipitation (figure 2 in ref. [46]). All are in the tree-ring chronology's lowest 37.1% of ARS values, with 6 of these 10 years in the lowest 12.9% of ARS values and 4 of the lowest 5 ARS values (7.1%) matching against these years (and the exception, 1933, immediately follows one of the top-10 driest years in both the meteorological records and ARS identified years, and so might be seen as related to the preceding very dry year). All but one of the lowest 20% of observed years (92.9%) are also in the lowest 37.1% of the ARS values. The single driest year, 1935, yields the second lowest ARS value. Thus, while in no way an entirely direct relationship, we might reasonably conclude that the occurrences of groups of values in the lowest quintile, and then lowest 10% of values, and especially the suggested lowest 6.25% of ARS values in the Gordion juniper series, and their comparison in the main text against the Polatlı meteorological data[34] probably (and qualitatively) indicates drier to very dry years. When identifying the lowest

6.25%, 10% and 20% of values, if more than one year had the threshold value, then the set included was reduced to be <6.25%, <10% and <20%.

Despite some variations, recent meteorological records demonstrate similar and largely coherent precipitation patterns for the Central Anatolia region (much of the core Hittite area) and for the south-central Black Sea region close/relevant to Hattusa[88,89,90,91], with all falling into the same Continental Central Anatolia (CCAN) precipitation region[88,89] (the approximate CCAN region is shown in Extended Data Fig. 1); the same precipitation shape classes, C or D, and magnitude classes, 1 (lowest precipitation values) or (one case only) 2 (second lowest values), and, overall, the inland composite precipitation regime regions as identified in ref. [90] (the combined approximate area of the two main composite inland precipitation regimes identified in ref. [90] at figure 8 are indicated in Extended Data Fig. 1). Instrumental (ad 1901–2012) and tree-ring derived Palmer Drought Series Index values ad 1500–2012[35,58], further attest to this general coherence across much of central Anatolia, and especially for years of marked low precipitation (Fig. 3 and Extended Data Figs. 6 and 7). Thus it is reasonable to assume that instances of severe/extreme drought in the Polatlı–Gordion area, as indicated by the tree-ring record, are likely to translate as arid to very arid conditions more widely across much of central Anatolia (whether the meteorologically broadly similar and coherent Continental Central Anatolia region[88,89], or the Central Anatolia and south-central Black Sea regions, or overall simplified inland precipitation regime regions, outlined in ref. [90]), that includes Polatlı–Gordion, Hattusa–Boğazkale, and the other loci employed in Extended Data Figs. 6 and 7. The data for grid points approximating Polatlı, Kirikkale, Boğazkale, Sivas and Kayseri in Extended Data Figs. 6 and 7 were extracted from ref. [58]. Hence, despite some intra-regional variation across distances of several hundred kilometers, we argue that use of the (unique) tree-ring series available from juniper trees that grew around the Gordion region (probably within ~30–50 km or so of the site, but perhaps even from a wider zone in central Anatolia[34,49,67]), before their use[34,49,66,67,68], provide insight into major drought events that widely affected much of central Anatolia, including the region around the Hittite capital of Hattusa, and the core Hittite region and subsistence base.

Figure 3 and Extended Data Figs. 6 and 7 use data from The Old World Drought Atlas[58] which are archived and freely available from the National Centers for Environmental Information (NCEI) at the National Oceanic and Atmospheric Administration (NOAA) (https://www.ncdc.noaa.gov/data-access/paleoclimatology-data). To select and represent these data we used the Tree-Ring Drought Atlas Portal[92] (freely available and open-access) located at http://drought.memphis.edu/ which facilitates analysis of the gridded reconstructions (last accessed November 2022).

Stable isotope data

The stable ^{13}C isotope ratio of plant material is a reflection of the ratio of leaf external to internal CO_2 concentrations (c_i/c_a) during growth. The c_i/c_a ratio is primarily controlled by stomatal conductance, the aperture of pores on the leaf surface allowing for the exchange of CO_2 and water vapour into and out of the leaf. During dry periods, plants close stomata to avoid water loss and hydraulic damage to plant tissue. Therefore, times of limited water availability are recorded in the cellulose of wood and, when combined with dendrochronologically dated tree rings, can be used as a first approximation of environmental drought history[93,94,95,96,97]. Previous isotope work on tree rings of *Juniperus* sp. and other conifer species in Anatolia identified a moisture availability signal[98,99]. To provide additional support for the drought hypothesis suggested in this study we investigated the $^{13}C/^{12}C$ isotope ratios of wood from four trees from the Gordion chronology: GOR-3 (two sections of the same tree but different parts and collected at different times), GOR-77, GOR-82 and GOR-87. This limited sample material was used to test whether there was a consistent indication of drier conditions around the period identified from the tree-ring width analysis, rather than production of a comprehensive, detailed, long-term stable ^{13}C based climatic record. Wood sections whose tree rings overlapped the dated time period of interest, had sufficient mass for isotope analysis and were most accessible for dissection, were dissected using a steel blade under a binocular microscope. Possible juvenile tree rings were avoided, with sampling starting from 50 years (GOR-82), 59 years (GOR-77), 113 years (GOR-3) to 383 years (GOR-87) after pith or start of sample series. Where possible (in sections with sufficiently wide ring widths and available

material) samples were dissected as single, annual, tree rings within the time period of the study. However, this was not always possible due to instances of very narrow tree rings and especially in cases of wedging tree rings or locally absent rings that reduced available material and practical sampling—in such cases the smallest number of practical tree rings was sampled. Whole annual ring samples were used (no division of early and late wood), whether for single tree rings where possible, or a multi-year group in other cases. For two series every second tree-ring was analysed. The samples were cut into small slivers with Soxhlet extraction used to remove labile carbon and then processed to α-cellulose in separate Teflon bags adapting a modified Jayme–Wise procedure[100].

The carbon isotope ratio ($\delta^{13}C$ versus Vienna Pee Dee Belemnite (VPDB)) data were obtained at the Cornell University Stable Isotope Laboratory using an isotope ratio mass spectrometer (Delta V, Thermo Scientific) interfaced with a high temperature pyrolysis analyser (TCEA, Thermo Scientific). The within-run isotopic precision of the methodology using quality control standards at the laboratory is ≤0.2‰ for carbon. The dataset employed from samples GOR-3S, 3E, 77A, 82A, and 87B is listed in Supplementary Table 2 (five sets, two from different samples from the same tree) and are shown in Extended Data Fig. 9. All data acquired are shown except for (1) the GOR-87 series was ended to exclude some data well after the time period of interest and (2) the GOR-87 series is shown from RY1287 as data were corrupted during the times preceding these data. In cases where a measurement was made on multiple tree rings, this value was applied to all of these tree rings; where data were collected on every second tree ring, linear extrapolation was used to create annual data for the comparisons reported. Since each individual tree derives from a unique microclimate, we do not expect them to have the same carbon isotope ratio. The exceptions are the two series from different elements of GOR-3. These show variation but a similar baseline level and both show rising levels relative year ~1230–1320 (~1282–1192 bc) to a consistent peak ~relative year 1316 (~1196 bc) (Extended Data Fig. 9). The important observation across the sets of values concern where there is a pattern of similar divergences in the trends recorded around the same times across several of the individuals, such as relative year ~1230–1320 (GOR-77A, 3E, 3S, 87B) (Extended Data Fig. 9). This strongly suggests each tree was experiencing

drier conditions, even though they have different baselines. There appears to be evidence of a small trend to less negative ^{13}C values with increasing tree age (Extended Data Fig. 9a, lower) consistent with observations of small age-related trends as recently quantified for a set of *Pinus sylvestris* L. trees from northern Fennoscandia[101]. However, the magnitude of this trend is much smaller than the diversions we have observed to indicate changes to drier conditions in this study. In particular, the key observation from Extended Data Fig. 9 is a general trend to less negative ^{13}C values in the period from around GOR RY1230 to 1320.

The ^{13}C data in Supplementary Table 2, and the different baselines and variations evident (the average value (±s.d.) from all data is −19.68‰ (±0.73‰)) correspond with other more recent time series of ^{13}C on juniper from Anatolia (for example, figure 2 in ref. [99]). The different baselines from different trees suggest that using and averaging the raw $\delta^{13}C$ values is not appropriate. Instead, since there appear to be indications of several common trends, we consider the data in terms of (normalized) *z*-transformed values for each individual $\delta^{13}C$ series (as in ref. [99]). The trajectories of the individual *z*-transformed $\delta^{13}C$ series exhibit a similar overall trend and were averaged into one mean site chronology, $\delta^{13}C_{CorZ}$. We then applied a 28-point Savitzky–Golnay (SG) smoothing curve to $\delta^{13}C_{CorZ}$ (with the 28-point (28-year) smoothing chosen as representative of average human generation timescale in pre-modern times[50]). We compare 28-point Savitzky–Golnay smoothing curves constructed from the individual data series versus this average (except in the case of GOR-82 where a 28-point adjacent average smoothing was applied as the data in this series are not all evenly spaced) (Extended Data Fig. 10). The smoothed individual series all correspond well with the smoothed $\delta^{13}C_{CorZ}$ curve (see Extended Data Fig. 10) offering correlation coefficients against the smoothed $\delta^{13}C_{CorZ}$ curve of GOR-3E (0.62), GOR-3S (0.9), GOR-77 (0.73), GOR-82 (0.80) and GOR-87 (0.78). The $\delta^{13}C_{CorZ}$ curve values from Extended Data Fig. 10 are thus employed as representative or indicative of general moisture availability in Fig. 2. We highlight that this analysis is presented in support of the ring-width analysis and not as a pure stand-alone climate proxy. The number of separate trees employed is low. The focus of our investigation is merely to

ascertain whether tree-ring stable carbon values indicate likely drier conditions (in relative terms irrespective of possible age-related and other trends) consistent with the tree-ring width analysis. The data and analysis we report indicate a relatively moderate (less arid) period from the mid-14th century bc to the earlier 13th century bc (green shaded areas) before a trend to noticeably drier conditions in the later 13th to initial 12th century bc (orange shaded areas) and then generally drier conditions from then until the late 12th century bc (Fig. 2). Outside the period of focus in this study, it is also noticeable that four trees indicate another arid episode in the record in the late-to-end 1020s bc (GOR RY1489–1494). There is a two-year consecutive lower growth interval GOR RY1487–1488, which may be associated (Fig. 1 and Extended Data Fig. 8a), but the tree-ring width response is neither as substantial nor clearly indicated as the GOR RY1314–1316 or ~1198–1196 bc case. The stable $\delta^{13}C$ data from Sofular Cave shown in Fig. 2 are from ref. [102]; the stable $\delta^{13}C$ and $\delta^{18}O$ data from Kocain Cave in Fig. 2 are from ref. [43]. While not of the same calendar-scale resolution, both these cave speleothem records offer complementary and consistent indications of drier conditions towards and around 1200 bc (as indicated by the arrows in Fig. 2).

Hittite and ancient Near Eastern history, geography and chronology

Hittite and surrounding ancient Near Eastern history and approximate geography are taken from refs. [4,11,12,13,14,15,21,23,24,25,26,29,32]. Additional sources, regarding Hittite history, Empire, economy and sociopolitical structures, as well as the environmental and landscape contexts ~1200 bc, include refs. [103,104,105,106,107,108,109,110]. The suggestion of drought as a, or the, possible cause of the Late Bronze Age collapse has a long history, first as a hypothesis; more recently it has been proposed as a century to multi-century-scale dry and cooler episode[1,2,3,4,30,31,111,112,113,114,115,116]. A variety of textual sources refer to grain shortages or to famine variously dated somewhere in the 13th century bc. Some more specifically date to the period of the late 13th century bc and around 1200 bc, and there are complementary indications of higher grain prices potentially consistent with shortages or supply problems[4,20,37,103,114,117,118]. What has been absent is

evidence of a specific but potentially historically transformative multi-year severe drought. However, at the same time, while many studies concentrate on the widespread evidence for collapse in the period around and following ~1200 bc[4,17,18,19,109,114], it is important to observe that, while widespread, collapse was by no means universal, and, at a number of sites (and areas), there is rather evidence of continuation, reorientation or even development during this time, and differential site and regional impacts and trajectories are observed[20,26,28,38,57,59,117,119,120,121,122]. In the case of the Hittites, in particular, it is the Hittite Empire and its central administration and the site of Hattusa (capital and religious centre or core), especially, that collapses (ending also the primary textual history available from the Hittite world until that time). Elements of the former Hittite world, smaller successor states in some cases, and some smaller sites in less challenged environmental settings even relatively proximate to Hattusa, like Çadır Höyük ~70 km distant, adapted and continued. Some other major sites, like Kaymakçı, show a marked decline from ~1200 bc, but perhaps include a final Late Bronze Age or Early Iron Age phase before abandonment[123]. As observed by ref. [122]: "depending on one's geographic location … the end of the Hittite Empire might have felt more like a bump or a bang". As noted earlier, 'collapse' is thus rather best viewed as a form of transformation and so resilience.

Dates for Hittite kings (for example, accession of Suppiluliuma II ~1207 bc) employ the (Mesopotamian) Middle Chronology (as ref. [11]). Both recent historical analysis, and dendrochronology combined with ^{14}C analysis, support a Middle Chronology date range for Mesopotamia (and thence associated groups like the Hittites)[69,124,125,126]. The terminus ante quem for the end, collapse and destruction of the Hittite Empire is the inscription at Medinet Habu in Egypt dated to the 8th year of Ramesses III, which lists *Hatti* (the Hittites) among those entities destroyed or swept away by the Sea Peoples[17,127]. This is assuming the strict historical accuracy of what is largely a highly political text presenting and celebrating the success and victory of Ramesses III, alone, in repelling what he describes as, until then, an overwhelming tide of change—the account contains several likely issues regarding its veracity when critiqued in detail[127]. In scholarship over the past several decades dates for the accession of Ramesses III were usually

placed in the mid-1180s bc (for example, 1187 bc[128] or 1184 bc[129]), and so his year 8 somewhere 1180–1177 bc, based around the fixed point of a date of 1279 bc for the accession of Ramesses II. However, recent work on the historical evidence[130,131,132], compatible with detailed radiocarbon assessment[133], instead suggests dates 11 years earlier: thus Ramesses III's year 8 would rather be placed at 1188 bc (and the accession of Ramesses II at 1290 bc). Other dates or last attested dates for Late Bronze Age sites in northern Syria that suffer destructions are in a similar range. For examples: (1) the destruction of the major coastal centre of Ugarit has been argued to lie between ~1193 and 1186 bc based on a synchronism with Bay, the Egyptian chancellor under Seti II but executed in the 5th year of Siptah—though these dates move ~11 years older if the earlier dates for Ramesses II and III noted above apply and strictly they form a terminus post quem rather than a date[20,118]; or (2) the site of Emar inland on the Euphrates was destroyed in the second year of the Babylonian king Melišipak (or Melišihu) ~1185 bc (Middle Chronology)[118,134,135,136]. The revised dates, with Ramesses III Year 8 placed ~1188 bc, and destructions, and mentions of grain shortages and famine, in a range from ~1200 to mid-1180s bc potentially offer a better fit with the dendro-^{14}C dated severe, multi-year drought ~1198–1196 ±3 bc—and thus the Middle Chronology dates for the last Hittite king Suppiluliuma II—than the alternative lower Egyptian or Mesopotamian dates. (The lower dates would place, for example, the collapse of the Hittites nearly 20 years after the serious drought, rather than, as the former, around or shortly after ~1198–1196 ± 3 bc).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

1. Kaniewski, D. et al. Cold and dry outbreaks in the eastern Mediterranean 3200 years ago. *Geology* **47**, 933–937 (2019).

2. Drake, B. L. The influence of climatic change on the Late Bronze Age collapse and the Greek Dark Ages. *J. Archaeolog. Sci.* **39**, 1862–1870 (2012).

3. Langgut, D. et al. Climate and the Late Bronze collapse: new evidence from the southern Levant. *Tel Aviv* **40**, 149–175 (2013).

4. Cline, E. H. *1177 B.C.: The Year Civilization Collapsed* (Princeton Univ. Press, 2021).

5. Halstead, P. & O'Shea, J. eds. *Bad Year Economics: Cultural Responses to Risk and Uncertainty* (Cambridge Univ. Press, 1989).

6. Butzer, K. W. Collapse, environment, and society. *Proc. Natl Acad. Sci. USA* **109**, 3632–3639 (2012).

7. Degroot, D. et al. Towards a rigorous understanding of societal responses to climate change. *Nature* **591**, 539–550 (2021).

8. Halstead, P. *Two Oxen Ahead: Pre-Mechanized Farming in the Mediterranean* (Wiley Blackwell, 2014).

9. Ó Gráda, C. Making famine history. *J. Econ. Lit.* **45**, 5–38 (2007).

10. Manning, S. W. in *Water and Power in Past Societies* (ed Holt, E.) 276–295 (State Univ. of New York Press, 2018).

11. Bryce, T. *The Kingdom of the Hittites* (Oxford Univ. Press, 2005).

12. Glatz, C. *The Making of Empire in Bronze Age Anatolia: Hittite Sovereign Practice, Resistance, and Negotiation* (Cambridge Univ. Press, 2020).

13. van den Hout, T. *A History of Hittite Literacy: Writing and Reading in Late Bronze Age Anatolia (1650–1200 BC)* (Cambridge Univ. Press, 2020).

14. Hoffner, H. A. Jr. *Letters from the Hittite Kingdom* (Society of Biblical Literature, 2009).

15. Murnane, W. J. *The Road to Kadesh: A Historical Interpretation of the Battle Reliefs of King Sety I at Karnak* 2nd edn (The Oriental Institute, Univ. of Chicago, 1990).

16. Middleton, G. D. *Understanding Collapse: Ancient History and Modern Myths* (Cambridge Univ. Press, 2017).

17. Sandars, N. K. *The Sea Peoples: Warriors of the Ancient Mediterranean 1250-1150 B.C.E.* (Thames & Hudson, 1987).

18. Ward, W. A. & Joukowsky, M. S. eds. *The Crisis Years: The 12th Century B.C.E. From Beyond the Danube to the Tigris* (Kendall/Hunt, 1992).

19. Bachhuber, C. & Roberts, R. G. eds. *Forces of Transformation: The End of the Bronze Age in the Mediterranean* (Oxbow Books, 2009).

20. Knapp, A. B. & Manning, S. W. Crisis in context: the end of the Late Bronze Age in the Eastern Mediterranean. *Am. J. Archaeol.* **120**, 99–149 (2016).

21. Seeher, J. in *Ipamati Kistamati Pari Tumatimis: Luwian and Hittite Studies Presented to J. David Hawkins on the Occasion of his 70th Birthday* (ed Singer, I.) 220–229 (Institute of Archaeology, Tel Aviv University, 2010).

22. Middleton, G. D. A tale of three cities: urban and cultural resilience and heritage between the Late Bronze Age and Early Iron Age in the Eastern Mediterranean. *Urban History* **48**, 724–748 (2021).

23. Schachner, A. in *Anatolia Between the 13th and 12th Century BCE* (ed. de Martino, S. & Devecchi, E.) 381–410 (LoGisma Editore, 2020).

24. Hawkins, J. D. in *Die Nahöstlichen Kulturen und Griechenland an der Wende vom 2. Zum 1. Jahrtausend v. Chr: Kontinuität und Wandel von Strukturen und Mechanismen Kultureller Interaktion: Kolloquium des Sonderforschungsbereiches 295 "Kulturelle und sprachliche Kontakte" der Johannes Gutenberg-Universität Mainz, 11–12. Dezember 1998* (eds Braun-Holzinger, E. A. & Matthäus, H.) 143–151 (Bibliopolis, 2002).

25. Bryce, T. *The World of The Neo-Hittite Kingdoms: A Political and Military History* (Oxford Univ. Press, 2012).

26. Osborne, J. F. *The Syro-Anatolian City-States: An Iron Age Culture* (Oxford Univ. Press, 2021).

27. Gilibert, A. *Syro-Hittite Monumental Art and the Archaeology of Performance* (De Gruyter, 2011).

28. Ross, J. C. et al. When the giant falls: endurance and adaptation at Çadır Höyük in the context of the Hittite Empire and its collapse. *J. Field Archaeol.* **44**, 19–39 (2019).

29. Weeden, M. After the Hittites: the kingdoms of Karkamish and Palistin in Northern Syria. *Bull. Inst. Class. Stud.* **56**, 1–20 (2013).

30. Kuzucuoğlu, C. in *La Cappadoce Méridionale de la Préhistoire à la Période Byzantine* (eds Beyer, D., Henry, O. & Tibet, A.) 17–41 (Institut Français d'Études Anatoliennes Georges-Dumézil, 2015).

31. Kaniewski, D. & Van Campo, E. in *Megadrought and Collapse: from Early Agriculture to Angkor* (ed. Weiss, H.) 161–182 (Oxford Univ. Press, 2017).

32. Bryce, T. The last days of Hattusa. *Archaeol. Od.* **8**, 32–41 (2005).

33. Köse, N. et al. Tree-ring reconstructions of May–June precipitation for western Anatolia. *Quat. Res.* **75**, 438–450 (2011).

34. Marston, J. M. *Agricultural Sustainability and Environmental Change at Ancient Gordion* (Univ. of Pennsylvania Museum of Archaeology and Anthropology, 2017).

35. Cook, B. I. et al. Spatiotemporal drought variability in the Mediterranean over the last 900 years. *J. Geophys. Res.* **121**, 2060–2074 (2016).

36. White, S. *The Climate of Rebellion in The Early Modern Ottoman Empire* (Cambridge Univ. Press, 2011).

37. Divon, S. A. in *The City of Emar Among the Late Bronze Age Empires: History, Landscape and Societies* (eds d'Alfonso, L. et al.) 101–109 (Ugarit, 2008).

38. Miller, J. in *Anatolia Between the 13th and the 12th Century BCE* (eds De Matino, S. & Devecchi, E.) 237–255 (LoGisma Editore, 2020).

39. Wright, N. J. Examining dendrological features of Oak as possible signals of systematic woodland management in the central Anatolian Bronze and Iron Ages. *Quat. Int.* **463**, 298–311 (2018).

40. Adcock, S. E. *After The End: Animal Economies, Collapse, and Continuity in Hittite and Post-Hittite Anatolia*. PhD dissertation, Univ. of Chicago (2020).

41. Diffey, C. et al. The agroecology of an early state: new results from Hattusha. *Antiquity* **94**, 1204–1223 (2020).

42. Scott, J. C. *Against the Grain* (Yale Univ. Press, 2017).

43. Jacobson, M. J. Heterogenous late Holocene climate in the Eastern Mediterranean—the Kocain Cave record from SW Turkey. *Geophys. Res. Lett.* **48**, e2021GL094733 (2021).

44. Garnsey, P. *Famine and Food Supply in the Graeco-Roman World* (Cambridge Univ. Press, 1988).

45. Xoplaki, E. et al. Modelling climate and societal resilience in the Eastern Mediterranean in the last millennium. *Hum. Ecol.* **46**, 363–379 (2018).

46. Touchan, R. et al. May–June precipitation reconstruction of southwestern Anatolia, Turkey during the last 900 years from tree rings. *Quat. Res.* **68**, 196–202 (2007).

47. Seim, A. et al. Climate change increases drought stress of juniper trees in the mountains of central Asia. *PLoS ONE* **11**, e0153888 (2016).

48. Kahveci, G. et al. Distribution of juniper stands and the impact of environmental parameters on growth in the drought-stressed forest-steppe zone of Central Anatolia. *Dendrobiology* **80**, 61–69 (2018).

49. Miller, N. F. *Botanical Aspects of Environment and Economy at Gordion, Turkey* (Univ. Pennsylvania Museum of Archaeology and Anthropology, 2010).

50. Moorjani, P. et al. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl Acad. Sci. USA* **113**, 5652–5657 (2016).

51. Sessa, K. The new environmental fall of Rome: a methodological consideration. *J. Late Antiq.* **12**, 211–255 (2019).

52. Kintigh, K. W. & Ingram, S. E. Was the drought really responsible? Assessing statistical relationships between climate extremes and cultural transitions. *J. Archaeolog. Sci.* **89**, 25–31 (2018).

53. Müller-Karpe, A. in *Tree-rings, Kings, and Old World Archaeology and Environment* (eds Manning, S. W. & Bruce, M. J.) 253–262 (Oxbow Books, 2009).

54. Harmanşah, Ö. in *New Materialisms: Ancient Urbanisms* (eds Alt, S. M. & Pauketat, T. R.) 218–244 (Routledge, 2020).

55. Raphael, S. K. *Climate and Political Climate: Environmental Disaster in the Medieval Levant* (Brill, 2013).

56. Wilkinson, T. C. *Tying the Threads of Eurasia: Trans-Regional Routes and Material Flows in Transcaucasia, Eastern Anatolia and Western Central Asia, c.3000-1500 BC* (Sidestone Press, 2014).

57. Manning, S. W. et al. Beyond megadrought and collapse in the Northern Levant: the chronology of Tell Tayinat and two historical inflection episodes, around 4.2ka bp, and Following 3.2ka bp. *PLoS ONE* **15**, e0240799 (2020).

58. Cook, E. R. et al. Old World megadroughts and pluvials during the Common Era. *Sci. Adv.* **1**, e1500561 (2015).

59. Broodbank, C. *The Making of the Middle Sea: A History of the Mediterranean from the Beginning to the Emergence of the Classical World* (Thames & Hudson, 2013).

60. Ulus, T. & Ellenblum, R. How long and how strong must a climatic anomaly be in order to evoke a social transformation? Historical and contemporaneous case studies. *Humanit. Soc. Sci. Commun.* **8**, 252 (2021).

61. Middleton, G. D. The show must go on: collapse, resilience, and transformation in 21st-century archaeology. *Rev. Anthropol.* **46**, 78–105 (2017).

62. Hegmon, M. et al. Social transformation and its human costs in the prehispanic U.S. Southwest. *Am. Anthropol.* **110**, 313–324 (2008).

63. Schweingruber, F. H. *Tree Rings: Basics and Applications of Dendrochronology* (D. Reidel, 1988).

64. Cook E. R. & Kairiukstis L. A. (eds) *Methods of Dendrochronology: Applications in the Environmental Sciences* (Kluwer, 1990).

65. Speer, J. H. *Fundamentals of Tree-Ring Research Tucson* (Univ. of Arizona Press, 2010).

66. Kuniholm, P. I. *Dendrochronology at Gordion and on the Anatolian Plateau.* PhD dissertation, Univ. of Pennsylvania (1977).

67. Kuniholm, P. I., Newton, M. W. & Liebhart, R. F. in *The New Chronology of Iron Age Gordion* (eds Rose, C. B. & Darbyshire, G.) 79–122 (Univ. of Pennsylvania Museum of Archaeology and Anthropology, 2011).

68. Manning, S. W. et al. ^{14}C record and wiggle-match placement for the Anatolian (Gordion Area) juniper tree-ring chronology ~1729 to 751 cal bc, and typical Aegean/Anatolian (growing season related) regional ^{14}C offset assessment. *Radiocarbon* **52**, 1571–1597 (2010).

69. Manning, S. W. et al. Radiocarbon offsets and Old World chronology as relevant to Mesopotamia, Egypt, Anatolia and Thera (Santorini). *Sci. Rep.* **10**, 13785 (2020).

70. Kealhofer, L. in *The Archaeology of Midas and the Phrygians: Recent Work at Gordion* (ed. Kealhofer, L.) 137–148 (Univ. of Pennsylvania Museum of Archaeology and Anthropology, 2005).

71. Marsh, B. in *The Archaeology of Midas and the Phrygians: Recent Work at Gordion* (ed. Kealhofer, L.) 161–171 (Univ. of Pennsylvania Museum of Archaeology and Anthropology, 2005).

72. Kahveci, G. & Arslan, M. Factors affecting the radial growth of *Juniperus foetidissima* Willd. and *J. excelsa* M. Bieb. in central Anatolia. *J. For. Sci.* **67**, 477–488 (2021).

73. Pearson, C. et al. Securing timelines in the ancient Mediterranean using multiproxy annual tree-ring data. *Proc. Natl Acad. Sci. USA* **117**, 8410–8415 (2020).

74. Holmes, R. L. Computer-assisted quality control in tree-ring dating and measurement. *Tree Ring Bull.* **43**, 69–78 (1983).

75. Baillie, M. G. L. & Pilcher, J. R. A simple crossdating program for tree-ring research. *Tree Ring Bull.* **33**, 7–14 (1973).

76. Fowler, A. M. & Bridge, M. C. Empirically-determined statistical significance of the Baillie and Pilcher (1973) *t* statistic for British Isles oak. *Dendrochronologia* **42**, 51–55 (2017).

77. Wigley, T. M. L. et al. Cross-dating methods in dendrochronology. *J. Archaeolog. Sci.* **14**, 51–64 (1984).

78. Sass-Klaassen, U. et al. Dendrochronological dating and provenancing of timber used as foundation piles under historic buildings in The Netherlands. *Int. Biodeterior. Biodegrad.* **61**, 96–105 (2008).

79. Cook, E. R. *A Time Series Analysis Approach to Tree Ring Standardization.* PhD Thesis, Univ. of Arizona (1985).

80. Cook, E. R. et al. Program ARSTAN (49v1b_MRWE), Autoregressive tree–ring standardization program (Tree–Ring Laboratory of Lamont–Doherty Earth Observatory, 2017).

81. Büntgen, U. et al. The influence of decision-making in tree ring-based climate reconstructions. *Nat. Commun.* **12**, 3411 (2021).

82. Melvin, T. M. et al. Time-varying-response smoothing. *Dendrochronologia* **25**, 65–69 (2007).

83. Kleese, S. Critical note on the application of the "two-third" spline. *Dendrochronologia* **65**, 125786 (2021).

84. Friedman, J. H. *A Variable Span Scatterplot Smoother* Technical Report No. 5 (Stanford University Laboratory for Computational Statistics, 1984).

85. Osborn, T. J., Biffa, K. R. & Jones, P. D. Adjusting variance for sample-size in tree-ring chronologies and other regional-mean

timeseries. *Dendrochronologia* **15**, 89–99 (1997).

86. Wigley, T. M. L. et al. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. *J. Clim. Appl. Meteorol.* **23**, 201–213 (1984).

87. Touchan, R. et al. Preliminary reconstructions of spring precipitation in southwestern Turkey from tree-ring width. *Int. J. Climatol.* **23**, 157–171 (2003).

88. Türkeş, M. Spatial and temporal analysis of annual rainfall variations in Turkey. *Int. J. Climatol.* **16**, 1057–1076 (1996).

89. Türkeş, M. & Erlat, E. Precipitation changes and variability in Turkey linked to the North Atlantic Oscillation during the period 1930–2000. *Int. J. Climatol.* **23**, 1771–1796 (2003).

90. Sariş, F., Hannah, D. M. & Eastwood, W. J. Spatial variability of precipitation regimes over Turkey. *Hydrol. Sci. J.* **55**, 234–249 (2010).

91. Altın, T. B. et al. Change in precipitation and temperature amounts over three decades in Central Anatolia, Turkey. *Atmos. Clim. Sci.* **2**, 107–125 (2012).

92. Burnette, D. J. The Tree-Ring Drought Atlas Portal: gridded drought reconstructions for the past 500–2,000 years. *Bull. Am. Meteorol. Soc.* **102**, 953–956 (2021).

93. McCarroll, D. & Loader, N. J. Stable isotopes in tree rings. *Quat. Sci. Rev.* **23**, 771–801 (2004).

94. Siegwolf, R. T. W., Brooks, J. R., Roden, J. & Saurer, M. (eds) *Stable Isotope in Tree Rings: Inferring Physiological, Climatic and Environmental Responses* (Springer, 2022).

95. Leavitt, S. W. Regional expression of the 1988 U.S. Midwest drought in seasonal $\delta^{13}C$ of tree rings. *J. Geophys. Res.* **112**, D06107 (2007).

96. Saurer, M. et al. The climate–carbon isotope relationship in tree rings and the significance of site conditions. *Tellus* **47B**, 320–330 (1995).

97. Gessler, A. et al. Stable isotopes in tree rings: towards a mechanistic understanding of isotope fractionation and mixing processes from the leaves to the wood. *Tree Physiol.* **34**, 796–818 (2014).

98. Mutlu, H. et al. Environmental and climatic signals from stable isotopes in Anatolia tree rings, Turkey. *Reg. Environ. Change* **12**, 559–570 (2012).

99. Heinrich, I. et al. Winter-to-spring temperature dynamics in Turkey derived from tree rings since ad 1125. *Clim. Dyn.* **41**, 1685–1701 (2013).

100. Loader, N. et al. An improved technique for the batch processing of small wholewood samples to α-cellulose. *Chem. Geol.* **136**, 313–317 (1997).

101. Torbenson, M. et al. Investigation of age trends in tree-ring stable carbon and oxygen isotopes from northern Fennoscandia over the past millennium. *Quat. Int.* **631**, 105–114 (2022).

102. Fleitmann, D. et al. Timing and climatic impact of Greenland interstadials recorded in stalagmites from northern Turkey. *Geophys. Res. Lett.* **36**, L19707 (2009).

103. Collins, B. J. *The Hittites and Their World* (Society of Biblical Literature, 2007).

104. Genz, H. in *The Philistines and Other 'Sea Peoples' in Text and Archaeology* (eds Killebrew, A. E. & Lehmann, G.) 469–478 (Society of Biblical Literature, 2013).

105. Beal, R. H. in *The Oxford Handbook of Ancient Anatolia* (eds Steadman, S. R. & McMahon, G.) 579–603 (Oxford Univ. Press, 2011).

106. De Martino, S. (ed) *Handbook Hittite Empire: Power Structures* (De Gruyter, 2022).

107. Dörfler, W. et al. in *Insights into Hittite History and Archaeology* (eds Genz, H. & Mielke, D. P.) 99–124 (Peeters, 2011).

108. Weeden, M. & Ullmann, L. Z. (eds) *Hittite Landscape and Geography* (Brill, 2017).

109. Drews, R. *The End of the Bronze Age: Changes in Warfare and the Catastrophe ca. 1200 B.C.* (Princeton Univ. Press, 1993).

110. Weeden, M. in *The Oxford History of the Ancient Near East* (eds Radner, K., Moeller, N. & Potts, D. T.) 529–622 (Oxford Univ. Press, 2022).

111. Carpenter, R. *Discontinuity in Greek Civilization* (Cambridge Univ. Press, 1966).

112. Bryson, R. A. et al. Drought and the decline of Mycenae. *Antiquity* **48**, 46–50 (1974).

113. Gorny, R. L. Environment, archaeology, and history in Hittite Anatolia. *Biblic. Archaeol.* **52**, 78–96 (1989).

114. Kaniewski, D. et al. Drought and societal collapse 3200 years ago in the Eastern Mediterranean: a review. *WIRES Clim. Change* **6**, 369–382 (2015).

115. Neugebauer, I. et al. Evidences for centennial dry periods at ~3300 and ~2800 cal. Yr bp from micro-facies analyses of the Dead Sea sediments. *Holocene* **25**, 1368–1371 (2015).

116. Kaniewski, D. et al. Climate change and social unrest: a 6,000-year chronicle from the Eastern Mediterranean. *Geophys. Res. Lett.* **47**, e2020GL087496 (2020).

117. Finkelstein, I. et al. Egyptian imperial economy in Canaan: reaction to the climate crisis at the end of the Late Bronze. *Age. Ägypten und*

Levante **27**, 249–260 (2017).

118. Singer, I. in *The Sea Peoples and Their World: A Reassessment* (ed Oren, E. D.) 21–33 (The University Museum, University of Pennsylvania, 2000).

119. Millek, J. M. in *The "Sea Peoples" Up-To-Date. New Research on the Crisis Years in the Eastern Mediterranean* (eds Fischer, P. M. & Burge, T.) 113–140 (Austrian Academy of Sciences, 2017).

120. Millek, J. M. Just how much was destroyed? The end of the Late Bronze Age in the Southern Levant. *Ugarit Forsch.* **49**, 239–274 (2018).

121. Millek, J. M. Destruction at the end of the Late Bronze Age in Syria: a reassessment. *Studia Eblaitica* **5**, 157–190 (2019).

122. Ross, J. C. et al. Anatolian empires: local experiences from Hittites to Phrygians at Çadır Höyük. *J. E. Med. Arch. Heritage Stud.* **7**, 299–320 (2019).

123. Roosevelt, C. H. et al. Exploring space, economy, and interregional interaction at a second-millennium B.C.E. citadel in central western Anatolia: 2014–2017 research at Kaymakçı. *Am. J. Archaeol.* **122**, 645–688 (2018).

124. Barjamovic, G. et al. *Ups and Downs at Kanesh–Observations on Chronology, History and Society in the Old Assyrian Period* (Nederlands Instituut voor het Nabije Oosten, 2012).

125. Manning, S. W. et al. Integrated tree-ring-radiocarbon high-resolution timeframe to resolve earlier second millennium bce Mesopotamian chronology. *PLoS ONE* **11**, e0157144 (2016).

126. Höflmayer, F. & Manning, S. W. A synchronized early Middle Bronze Age chronology for Egypt, the Levant, and Mesopotamia. *J. Nr E. Stud.* **81**, 1–24 (2022).

127. Redford, D. B. in *The Sea Peoples and Their World: A Reassessment* (ed Oren, E. D.) 1–20 (The Univ. Museum, Univ. of Pennsylvania, 2000).

128. Hornung, E. et al. (eds) *Ancient Egyptian Chronology* (Brill, 2006).

129. Kitchen, K. A. In *The Synchronisation of Civilisations in the Eastern Mediterranean in the Second Millennium B.C.* (ed. Bietak, M.) 39–52 (Verlag der Österreichischen Akademie der Wissenschaften, 2000).

130. Schneider, T. Contributions to the chronology of the New Kingdom and the Third Intermediate period. *Ägypten Levante* **20**, 373–403 (2010).

131. Aston, D. Radiocarbon, wine jars and New Kingdom chronology. *Ägypten Levante* **22**, 289–315 (2012-2013).

132. Miller, J. L. in *Karduniaš. Babylonia Under the Kassites* (eds Bartelmus, A. & Sternitzke, K.) 93–111 (De Gruyter, 2017).

133. Bronk Ramsey, C. et al. Radiocarbon-based chronology for Dynastic Egypt. *Science* **328**, 1554–1557 (2010).

134. Arnaud, D. Les textes d'Emar et la chronologie de la fin du Bronze recent. *Syria* **52**, 87–92 (1975).

135. Brinkman, J. A. *Materials and Studies for Kassite History 1. A Catalogue of Cuneiform Sources Pertaining to Specific Monarchs of the Kassite Dynasty* (Oriental Institute of the Univ. of Chicago, 1976).

136. Brinkman, J. A. in *Karduniaš. Babylonia Under the Kassites* (eds Bartelmus, A. & Sternitzke, K.) 1–44 (De Gruyter, 2017).

137. Touchan, R. et al. Standardized precipitation index reconstructed from Turkish tree-ring widths. *Clim. Change* **72**, 339–353 (2005).

Acknowledgements

This research was partially supported by the CRANE project (award 895-2011-1026, Social Sciences and Humanities Research Council, Canada) (https://www.crane.utoronto.ca/index.html). We thank past members of the Cornell Tree-Ring Laboratory who have contributed to work on Gordion samples; and K. Sparks and the staff of the Cornell Stable Isotope Laboratory for their assistance. For support towards open-access publication, we thank the Department of Classics, the College of Arts & Sciences, the Society for the Humanities and the Cornell Stable Isotope Laboratory, Cornell University and the CRANE project.

Author information

Authors and Affiliations

1. Cornell Tree-Ring Laboratory, Department of Classics and Cornell Institute of Archaeology and Material Studies, Cornell University, Ithaca, NY, USA

   Sturt W. Manning & Brita Lorentzen

2. The Cyprus Institute, Nicosia, Cyprus

   Sturt W. Manning

3. Mississippi Valley Archaeology Center, University of Wisconsin-La Crosse, La Crosse, WI, USA

   Cindy Kocik

4. Department of Anthropology, University of Georgia, Athens, GA, USA

   Brita Lorentzen

5. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

   Jed P. Sparks

Contributions

S.W.M. conceived the idea, directed the research, performed the analysis and wrote the paper. C.K. carried out the majority of the sample preparation and laboratory stable isotope analysis. B.L. contributed to the dendrochronological and historical analyses. J.P.S. advised on and supervised the stable isotope analysis. All co-authors contributed to the final draft of the manuscript.

Corresponding author

Correspondence to Sturt W. Manning.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Dagomar Degroot, Ömür Harmansah, Jennifer Ramsay and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Map of East Mediterranean-Near East indicating the major ancient empires (Hatti = Hittites, Egypt, Mitanni, Assyria, Babylonia), several other main political entities sometimes within, sometimes on the margins of, the Hittite Empire (Kaska, Wilusa, Seha River Land, Arzawa-Mira, Lukka, Pitassa, Tarhuntassa, Alashiya, Kizzuwatna)11,12, the main sites/locations mentioned in this paper (Hattusa next to modern Boğazkale, Gordion, Polatlı, Sofular Cave, Kocain Cave, Kirikkale, Sivas, Kayseri, Karkemish, Kadesh), and the approximate Continental Central Anatolia (CCAN) precipitation zone[88,89] or the combined two

main simplified composite 'inland' precipitation regimes identified for Anatolia in ref. [90].

Note: Wiyanawanda, Masa, and Ikkuna, mentioned in the main text, are all thought to lie in, or close to, the Lukka territory in southwest Anatolia[11]. Map created from free and open (public domain) Natural Earth data (https://www.naturalearthdata.com/), using the Natural Earth II dataset.

Extended Data Fig. 2 Gordion tree-rings series and crossdating grid.

a, Raw (non-standardized) measured tree-ring series for the Gordion samples (all same scaling on Y axis) indicating the overall growth trends in each case. **b**, Crossdating grid for the Gordion tree-ring samples listing the Baillie-Pilcher t-scores (upper value) and length of overlap in rings/years (lower value) for each sample series versus the others from COFECHA[74]. Potentially satisfactory values where there is strong overlap (e.g. > 50-100 years) are $t_{BP} \geq 3.5$ (unsatisfactory values are indicated by grey shading)—although in reality higher t_{BP}-values are necessary for a secure match (see Methods). Higher values and especially t_{BP}-scores >5-6 usually indicate a good crossdate. Here all samples exhibit at least several strong to very strong crossdates against a number of other samples in the chronology with long overlaps, demonstrating an overall robust chronology.

Extended Data Fig. 3 Ring-width measurements for the Gordion chronology and processing with ARSTAN[79,80] and detrending to Standard Chronologies for the six different approaches used (see Fig. 1, Extended Data Figs. 5, 8a), each offering good correspondence with the data.

Sample depth is shown under each figure element.

Extended Data Fig. 4 Residual and ARSTAN (ARS) chronologies derived from the six Standard Chronologies in

Extended Data Fig. 3 using ARSTAN[79,80].

Sample depth is shown under each figure element.

Extended Data Fig. 5 Sample number per year and Expressed Population Signal86 for a set of six ARSTAN (ARS) chronologies derived from the Gordion samples (see Extended Data Figs. 3, 4) using the ARSTAN software[79,80] employing different standardization curves, each aimed at highlighting the likely underlying climate signal (see Methods).

The shaded region indicates where the EPS values are ≥0.85. Note for the Hugershoff detrending one sample (GOR-3) receives the ‘itmax exceeded in amoeba’ warning—we used the age dependent spline for this sample.

Extended Data Fig. 6 Modern Instrument-derived self-calibrating Palmer Drought Series Index (scPDSI) values and correlations ad 1901–2012 for January-August for central Anatolia from the Old World Drought Atlas (OWDA)58.

a, Comparison of a west-east transect across the core Hittite region in central Anatolia from Polatlı (39.58N, 32.14E) to Kirikkale (39.85N, 33.5E) to Boğazkale (40.02N, 34.61E) (location of Hattusa) to Sivas (39.75N, 37.02E) and for a major center to the southeast (Kayseri) (38.72N, 35.48E) (see Extended Data Fig. 1). The scPDSI series from Polatlı has Pearson correlation coefficients (r) = 0.78, 0.7, 0.48, and 0.6 respectively; thus positive and suggesting a general regional pattern, except for Sivas in the east. **b**, Critically, the one extreme drought event in the dataset, centered 1928, occurs as a < scPDSI -4 event across all five loci, and **c**., is reconstructed across the entirety of central Anatolia, suggesting that extreme arid events recorded at Polatlı are likely regionally effective (see also Fig. 3 and Extended Data Fig. 7).

Extended Data Fig. 7 Tree-ring derived reconstructed self-calibrating Palmer Drought Series Index (scPDSI) values and correlations ad 1500–2000 (when there is reasonable data representation for Anatolia35) for June-July-August for central Anatolia from the Old World Drought Atlas (OWDA)[58].

a, Comparison of values from a west-east transect across the core Hittite region in central Anatolia as in Extended Data Fig. 6. Pearson correlation coefficients (*r*) with Polatlı (in vertical order shown) = 0.86, 0.85, 0.89 and 0.89 indicating a strong regionally coherent pattern. Instances where Polatlı and either 3 (of 4) or all 4 other locations have scPDSI less than −2.7 (value indicating lowest 6.25% of values at Polatlı) are indicated, as are some historically attested or reconstructed droughts starting ad 1567 (ref. [46] at Table 3), the reconstructed 2yr dry episodes for western Anatolia ad 1786-1930 (W dry 2yr) (ref. [33] at Table 6) and instances of noted occurrences of 2yrs or 5yrs of drought in southern Anatolia (S dry 2/5yr) since 1500[137]. All the five OWDA PDSI series record the same major arid episodes with **b**, showing that there is a strong wider regional expression across central Anatolia for the extreme arid events such as the five driest years in the Polatlı series in the OWDA dataset. In contrast, some of the other historically recorded droughts or those reconstructed for western or southern Anatolia do not correlate with especially dry conditions in one or more of the Polatlı, Kirikkale, Boğazkale, Sivas, and Kayseri series—events where the dashed grey bars do not correspond with the magenta lines in **a**—and so may not reflect serious drought expressed broadly across central Anatolia.

Extended Data Fig. 8 Additional ARSTAN (ARS) climate (moisture availability) proxies and comparison of ARS index values for a SW Anatolia juniper chronology versus observed low precipitation years.

a, Three additional alternative versions of the Fig. 1 analyses. Drier calculated conditions (low index values) toward the bottom (compare with Fig. 1). See Fig. 1 for description of analysis. Grey bar indicates 12-year

period ~1198–1187 bc with 3 consecutive years ~1198–1196 in lowest 6.25% of all years and with 7 or 8 years (58–67%) altogether in lowest 20% of values. **b**, ARS rt-gt values for three juniper tree-ring chronologies from SW Anatolia[87], ad 1931–2000, versus the lowest observed precipitation totals for May-June, the period critical to annual precipitation response, from proximate meteorological stations[46]. The lowest (minimal) 37.1% of ARS values include the lowest 20% of observed precipitation values. The lowest 20% of ARS values include 50% of the lowest 20% of observed values and 7 of 9 (78%) in the period ad 1931–1961. The lowest 10% and 6.25% of ARS values are also indicated; with 5 of 7 (71%) of the former and 3 of 4 (75%) of the latter corresponding with years in the lowest 14.3% of observed years.

Extended Data Fig. 9 Gordion δ^{13}C data, raw and residual values, and trends.

a, Upper: plots of the δ^{13}C data (with linear extrapolation between data points). The linear fit lines indicate small gradients towards less negative values with increasing tree age[101]. The arrows indicate a marked trend towards less negative values from ~GOR RY1230 to 1320. Lower: residual plots (from linear fits) removing different baselines and the age-trends. Gradient ~GOR RY 1230–1320 is evident. **b**, δ^{13}C values from the two series on tree GOR-3. Similar baselines, and, despite noise, similar trend GOR RY1220–1320 (indicated by arrow). **c**, Average residual values (from **a**, lower) with 28 and 56 pt Savitzky-Golnay (SG) smoothing lines (one and two human generations[50]) for period mainly with ≥2 data per year. Trend to less negative values GOR RY1220–1320 is indicated (see also inset). The plot at the bottom indicates the number of data values per year.

Extended Data Fig. 10 Indicative δ^{13}C data and analysis from four Gordion trees (see SI, Table 2).

Z scores for the five δ^{13}C series comparing these via 28 pt Savitzky-Golnay (SG) smoothings (or 28 pt adjacent average smoothing in the case where the data are not all evenly spaced) versus the 28 pt SG smoothed curve of

the average z values for the combined $\delta^{13}C$ series ($\delta^{13}C_{CorZ}$) (see Methods). The correlation coefficients between compared smoothed curves versus the smoothed average curve are GOR-3E (0.62), GOR-3S (0.9), GOR-77 (0.73), GOR-82 (0.80) and GOR-87 (0.78).

Supplementary information

Supplementary Tables

This file contains Supplementary Tables 1 and 2

Reporting Summary

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Manning, S.W., Kocik, C., Lorentzen, B. *et al.* Severe multi-year drought coincident with Hittite collapse around 1198–1196 bc. *Nature* **614**, 719–724 (2023). https://doi.org/10.1038/s41586-022-05693-y

- Received: 14 March 2022
- Accepted: 30 December 2022
- Published: 08 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05693-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Tree rings reveal drought that might have doomed the Hittites**
  - Miryam Naddaf

  Nature (2023)

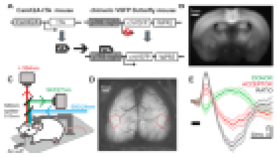
Signs of ancient climate crisis as the Hittite empire unravelled

- Müge Durusu-Tanrıöver

Nature News & Views 08 Feb 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05693-y

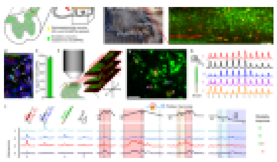
| Section menu | Main menu |



Cortical signatures of wakeful somatosensory processing

10 August 2018

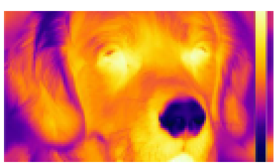
Chenchen Song, Denise M. Piscopo, … Thomas Knöpfel



Cell type-specific calcium imaging of central sensitization in mouse dorsal horn

03 September 2022

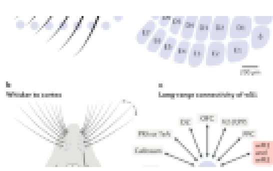
Charles Warwick, Joseph Salsovic, … Sarah E. Ross



Dogs can sense weak thermal radiation

28 February 2020

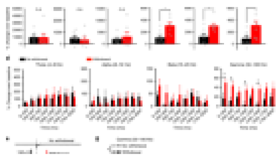
Anna Bálint, Attila Andics, … Ronald H. H. Kröger



[Sensorimotor processing in the rodent barrel cortex](#)

31 July 2019

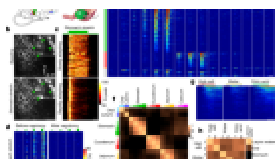
Carl C. H. Petersen



[Gamma oscillations in somatosensory cortex recruit prefrontal and descending serotonergic pathways in aversion and nociception](#)

28 February 2019

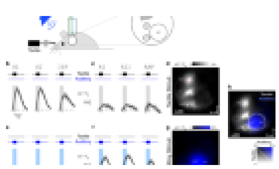
Linette Liqi Tan, Manfred Josef Oswald, … Rohini Kuner



[A brainstem map for visceral sensations](#)

31 August 2022

Chen Ran, Jack C. Boettcher, … Stephen D. Liberles



[Spectral hallmark of auditory-tactile interactions in the mouse somatosensory cortex](#)

11 February 2020

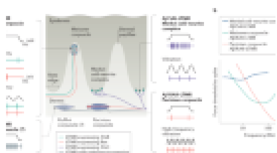
Manning Zhang, Sung Eun Kwon, … John B. Issa



[**Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types**]

23 April 2018

Martin Häring, Amit Zeisel, … Patrik Ernfors



[**The mechanosensory neurons of touch and their mechanisms of activation**]

26 July 2021

Annie Handler & David D. Ginty

- Article
- Open Access
- Published: 08 February 2023

The cellular coding of temperature in the mammalian cortex

- M. Vestergaard ,
- M. Carta ,
- G. Güney &

- …
- J. F. A. Poulet

Show authors

Nature **volume 614**, pages 725–731 (2023)

- 10k Accesses
- 161 Altmetric
- Metrics details

Abstract

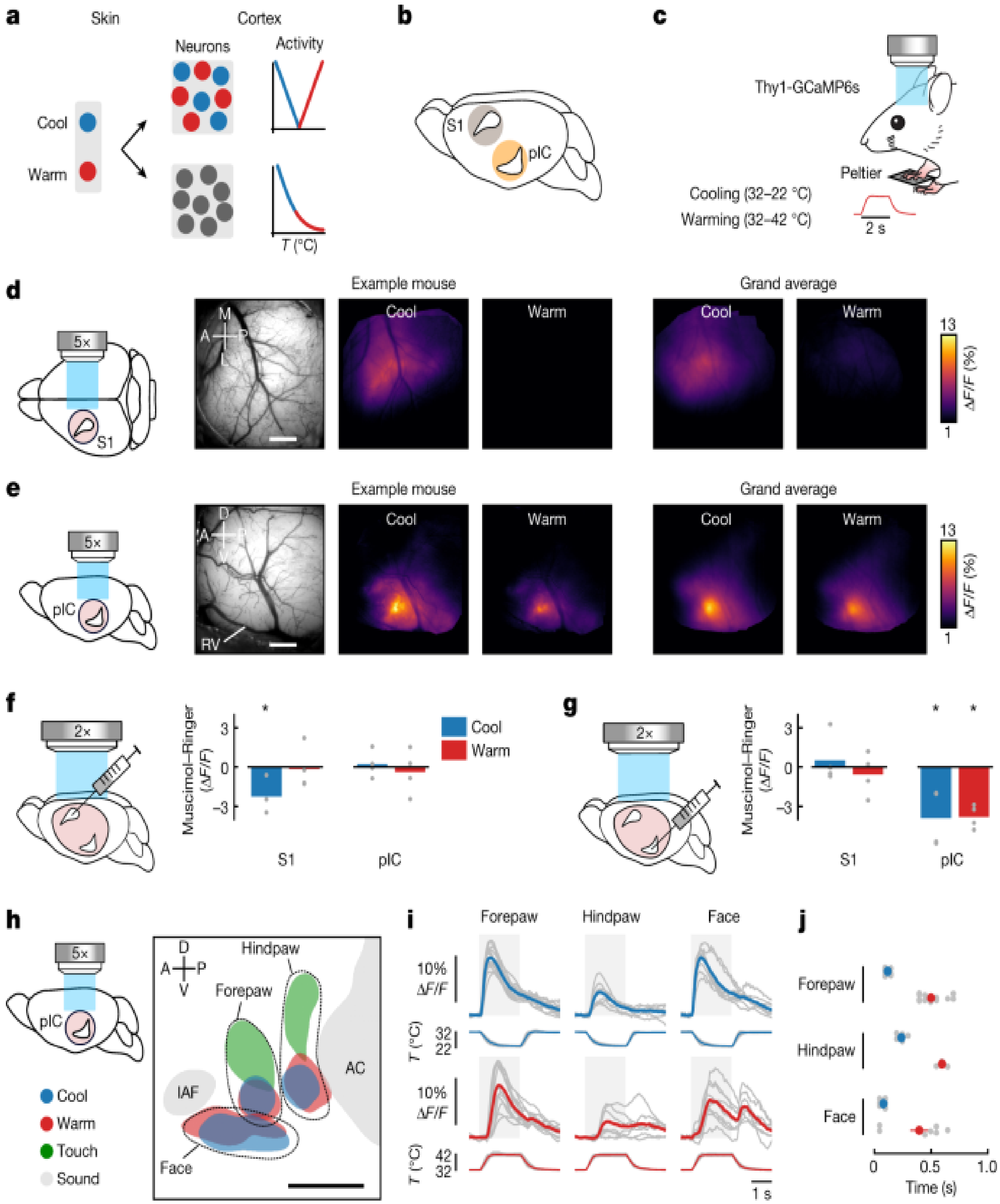
Temperature is a fundamental sensory modality separate from touch, with dedicated receptor channels and primary afferent neurons for cool and warm[1,2,3]. Unlike for other modalities, however, the cortical encoding of temperature remains unknown, with very few cortical neurons reported that respond to non-painful temperature, and the presence of a 'thermal cortex' is debated[4,5,6,7,8]. Here, using widefield and two-photon calcium imaging in the mouse forepaw system, we identify cortical neurons that respond to cooling and/or warming with distinct spatial and temporal response properties. We observed a representation of cool, but not warm, in the primary somatosensory cortex, but cool and warm in the posterior insular cortex (pIC). The representation of thermal information in pIC is robust and somatotopically arranged, and reversible manipulations show a profound impact on thermal perception. Despite being positioned along the same one-dimensional sensory axis, the encoding of cool and that of warm are distinct, both in highly and broadly tuned neurons. Together, our results show that pIC contains the primary cortical representation of skin temperature and may help explain how the thermal system generates sensations of cool and warm.

Main

A fundamental question in neuroscience is how the external sensory environment is represented in the cortex. In the thermal system, at present

there is no consensus on how or where sensory information is encoded in the cortex. One model is that cool and warm are processed by functionally and anatomically segregated circuits, following labelled-line principles seen in primary sensory afferent neurons and spinal circuits[9,10,11,12,13,14,15,16,17] resulting in cool- and warm-selective cortical neurons (Fig. 1a, top). Another is that cool and warm are integrated in the thermal pathway resulting in cortical neurons with a continuous, graded representation of temperature (Fig. 1a, bottom). Testing these models requires the identification of thermally responsive neurons in the cortex.

Fig. 1: Cortical representation of cool and warm.



a, Schematic showing segregated (top) and integrated (bottom) models of cortical thermal encoding. *T*, temperature. **b**, Mouse brain showing candidate locations of thermal cortex in primary somatosensory cortex (S1) or pIC. **c**, Schematic of an awake Thy1-GCaMP6s mouse with right forepaw on a Peltier element during widefield calcium imaging; inset shows temporal

dynamics of warming stimulus. **d**, From left to right: schematic of glass window on S1 (pink circle); in vivo image of cortical surface; averaged widefield response to 10 °C cooling (32–22 °C) or warming (32–42 °C) of the forepaw in an example mouse; grand average across mice ($n = 348$ cool and 346 warm trials, 12 mice). M, medial; A, anterior; L, lateral; P, posterior. Scale bar, 500 µm. **e**, Same as **d**, but for mice with glass window implanted over pIC ($n = 360$ cool and 352 warm trials, 14 mice). RV, rhinal vein; D, dorsal; V, ventral. Scale bar, 500 µm. **f**, From left to right: schematic illustrating injection site into S1 during widefield imaging through a clear skull preparation with simultaneous imaging of S1 and pIC (pink shows field of view); bar graphs show difference in response amplitude following muscimol versus Ringer's injection. Bars indicate mean and grey filled circles indicate individual mice ($n = 4$). An asterisk indicates significant difference in S1 response following muscimol versus Ringer's injection ($P < 0.05$, two-sided paired *t*-test, see Methods for exact values). **g**, Same as **f** but showing a reduction in pIC response to thermal stimuli after pIC inactivation and no change in the S1 response ($n = 4$ mice). **h**, Somatotopic map of response locations to thermal (10 °C cool and warm), tactile (100 Hz) and acoustic (8 kHz) stimulation. Coloured area indicates peak population response averaged across mice (see Methods, $n = 14$ mice thermal forepaw, 7 thermal hindpaw, 9 thermal face, 9 touch forepaw, 6 touch hindpaw, 13 sound). Data from individual mice are aligned to peak activity of the thermal forepaw response. IAF, insular auditory field; AC, auditory cortex. Scale bar, 500 µm. **i**, Widefield responses to thermal stimulation of different body parts from the same dataset as in **h**. Grey lines show mean responses from individual mice (*n* is same as in **h**), coloured lines show population mean, and grey area indicates time from start of stimulus to end of plateau phase. **j**, Grey filled circles show response latencies from individual mice (*n* is same as in **h**, see Methods); coloured filled circles show mean ± s.e.m.

Source data

Psychophysical studies have shown that cooling and warming elicit distinct sensations, influenced by stimulus amplitude, duration, dynamics, body part and adapted skin temperature[2,18]. Analogously to other modalities, these features should be represented in a cortical region dedicated to thermal processing. Moreover, manipulation of its activity should influence thermal

perception. A number of cortical regions have been suggested to be involved in thermal processing, including the primary somatosensory cortex (S1)[6,7,8] and the pIC[5,19,20] (Fig. 1b). However, no study has identified a cortical area with a cellular representation of cool and warm and a reversible impact on perception.

pIC contains a somatotopic map of cool and warm

To examine thermal processing in S1 and pIC, we carried out widefield calcium imaging through a glass window in paw-tethered, awake mice expressing the calcium indicator GCaMP6s in cortical excitatory neurons (Fig. 1b,c and Extended Data Fig. 1). We presented an 8-kHz acoustic stimulus to locate the anterior auditory field in the auditory cortex and the small auditory insular field that border the somatosensory regions of pIC[21,22,23], and went on to carry out post hoc histology (Extended Data Fig. 1). Next, we delivered 2-s thermal stimuli to the forepaw glabrous skin through a Peltier element held at an adapted temperature (AT) of 32 °C. Whereas 10 °C (32 °C to 22 °C) cooling stimuli triggered reliable changes in fluorescence in S1, 10 °C warming (32 °C to 42 °C) did not (Fig. 1d). By contrast, both cooling and warming stimuli triggered robust, large-amplitude responses in pIC (Fig. 1e). Local pharmacological inactivation of either S1 or pIC during imaging abolished the thermal response in the injected region but not in the untreated region (Fig. 1f,g), indicating that pIC and S1 receive parallel streams of thermal input.

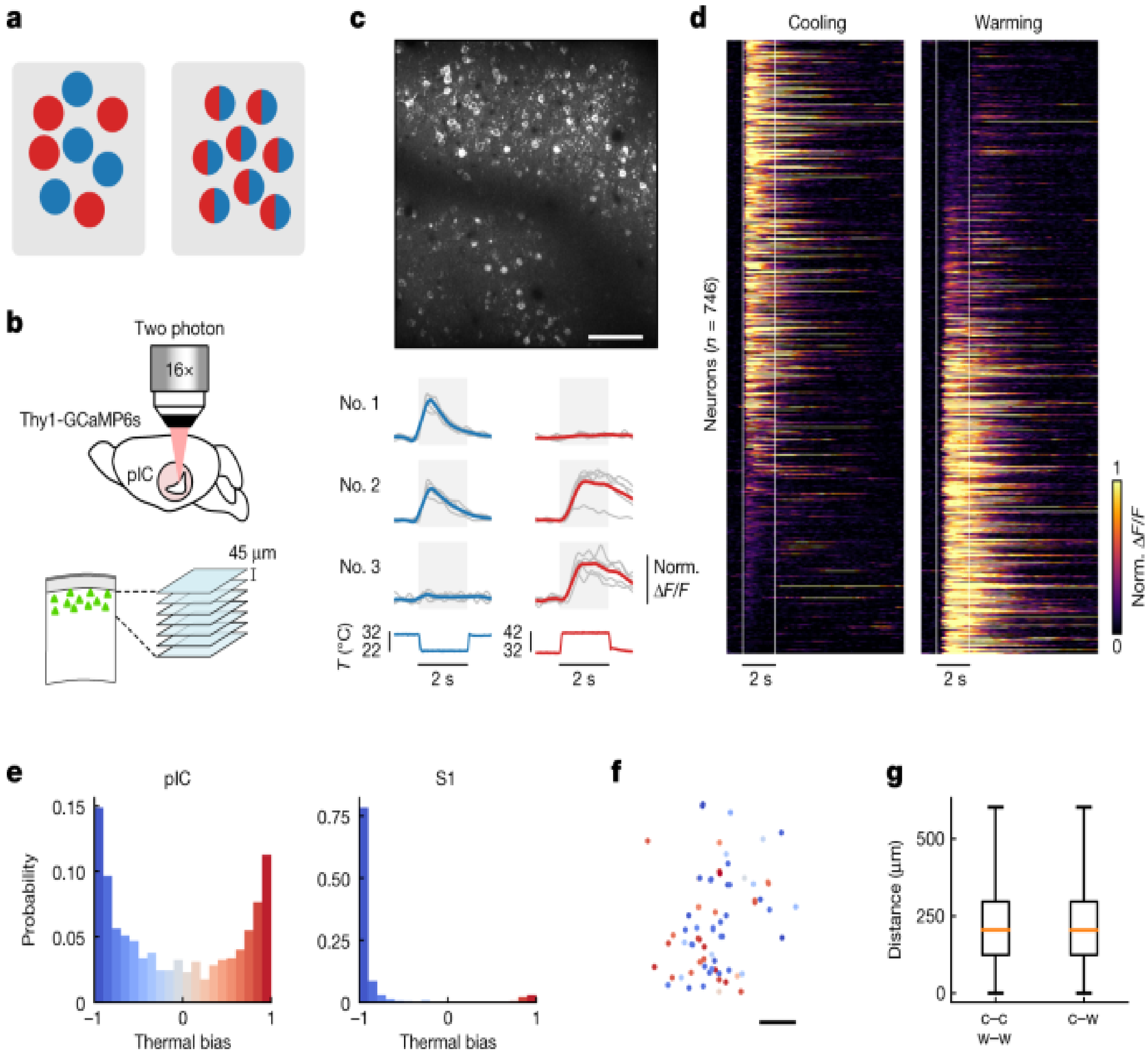
Perceptual thresholds for cool and warm covary across the body[24], suggesting that some body parts have a stronger cortical representation than others. Widefield calcium imaging of pIC showed a clear somatotopic arrangement of thermal and tactile responses anterior to the auditory cortex in all mice (Fig. 1h) with the face represented in a region anterior and ventral to the forepaw, which in turn is anterior to the hindpaw (Fig. 1h). In all mice, the forepaw had a larger-amplitude response to thermal stimuli than the hindpaw, perhaps reflecting the dominant role of the forepaw in haptic exploration (Fig. 1i). Whereas responses to cooling and touch spatially overlap in S1 (Extended Data Fig. 2; ref. [7]), they seem more separate in pIC (Fig. 1h). For all body parts tested, warming responses in pIC were delayed

compared to cooling responses, suggesting that neural response latency is a hardwired property of the thermal system (Fig. 1j).

Thermal tuning and topography of pIC neurons

Cooling and warming widefield responses in pIC overlap spatially (Fig. 1e,h). This could result from an intermingled distribution of highly tuned or broadly tuned neurons (Fig. 2a). To determine the thermal tuning of cortical neurons, we went on to carry out two-photon calcium imaging of pIC excitatory neurons in awake mice and observed robust cellular responses to cooling and warming (Fig. 2b–d). We went on to plot the normalized responses according to a thermal bias index that describes the relative response strength to cooling compared to warming, with a value of −1 indicating a cool-only cell and +1 indicating a warm-only cell (Fig. 2d,e). The index had a U-shaped distribution, suggesting that segregated channels of tuned input drive cool and warm responses, and showed a similar probability of highly tuned neurons (cool only or warm only) and broadly tuned neurons (cool and warm responsive). In line with widefield data, most neurons in S1 were tuned to cooling only and only a tiny fraction responded to warming with delayed and inconsistent responses (Figs. 2e and 3c and Extended Data Fig. 3).

Fig. 2: Heterogeneous arrangement of thermally tuned neurons in the pIC.

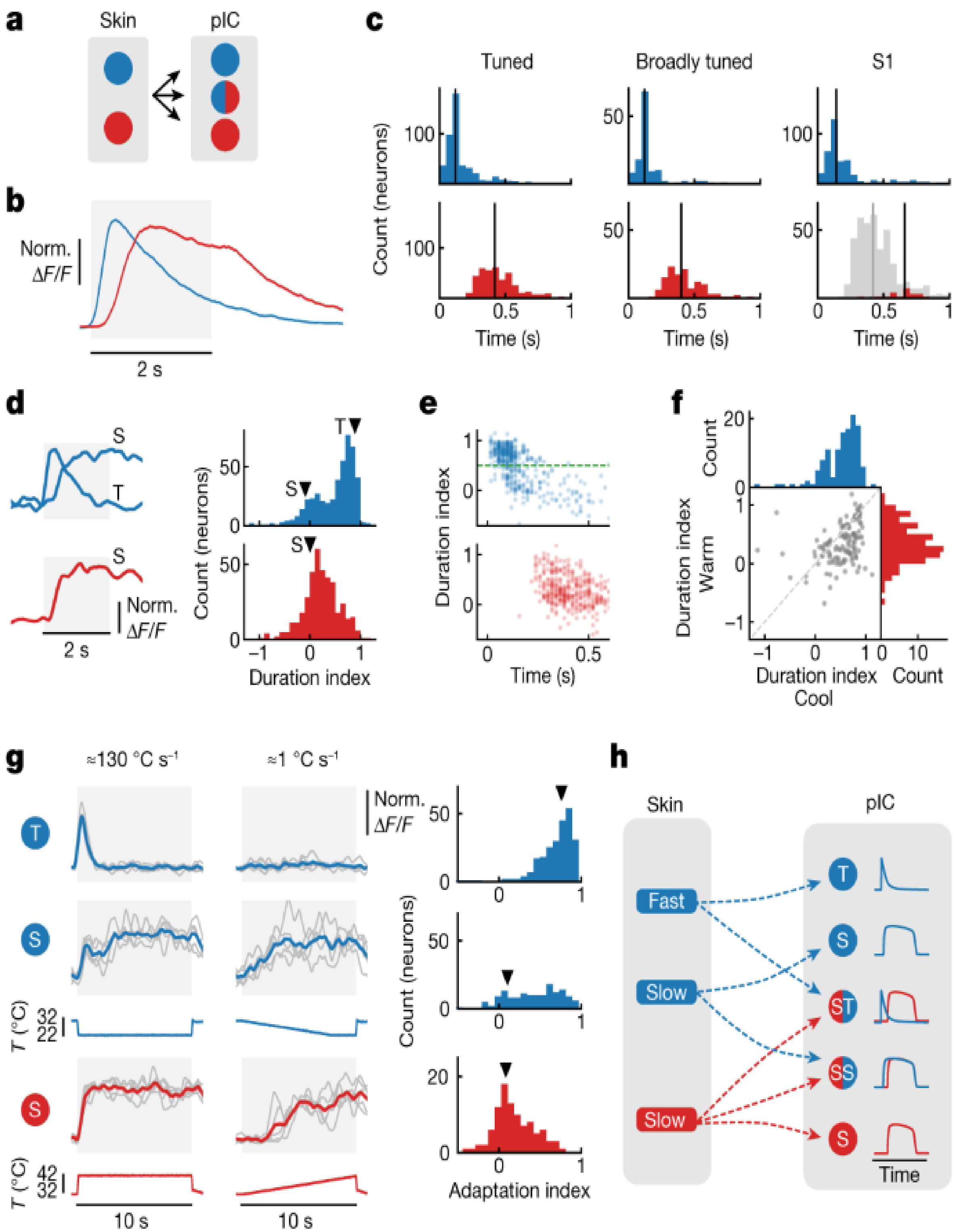


a, Schematic showing tuned (left) versus broadly tuned (right) cortical neurons. Blue indicates cool; red indicates warm. **b**, Schematic showing two-photon calcium imaging of pIC. Imaging started at 100 µm from the pial surface of a Thy1-GCaMP6s mouse; 7 optical sections were acquired with intervals of 45 µm. Norm., normalized. **c**, Top, example in vivo two-photon image from pIC. Scale bar, 100 µm. Bottom, example responses of single pIC neurons during 10 °C cooling (blue) or 10 °C warming (red) stimuli from an AT of 32 °C ($n = 5$ trials each). Here, and for all two-photon data figures, grey lines show single-trial responses, coloured lines show average, and background grey box indicates time from start of stimulus to end of plateau phase. Below, corresponding stimulus traces. **d**, Single-cell pIC calcium responses to cooling (left) and warming (right) stimuli. Each line represents a single neuron; responses are normalized to the peak and sorted

on the basis of the thermal bias index with vertical white lines showing the onset and the end of the plateau phase of thermal stimuli (*n* = 746 neurons, 7 mice, 16 sessions). **e**, Histograms showing the distribution of thermal bias for all responsive cells in pIC (left; *n* is same as in **d**) and S1 (right; *n* = 411 neurons, 4 mice, 9 sessions). **f**, Spatial map of neurons colour coded with their thermal bias for the representative mouse shown in **c**. Scale bar, 100 µm. **g**, No significant difference in the distance between pairs of neurons with similar bias (cold to cold (c–c) or warm to warm (w–w)) and distance between pairs of neurons with opposite bias (cold to warm (c–w)) indicates no thermotopy for strongly biased neurons (*n* is same as in **d**) (difference of medians confidence interval (−2 µm, 5 µm), 95%, bootstrapped). Box plots show median and interquartile range. Whiskers show minimum and maximum values.

Source data

Fig. 3: Distinct temporal dynamics of cool and warm encoding.



a, Schematic showing transition from cool- and warm-selective skin spots to tuned and broadly tuned pIC neurons. Blue indicates cool; red indicates warm. **b**, Grand average responses (mean ± s.e.m.) to 10 °C cooling and

warming responses from AT of 32 °C ($n = 470$ cool neurons (blue line), 401 warm neurons (red line) (see Methods); these neurons are also used in **c**–**f**). **c**, Histograms of the response latency for 10 °C cooling (top) and 10 °C warming (bottom), in (left to right) tuned neurons (n is as in **b**), broadly tuned neurons ($n = 125$ neurons) and S1 neurons ($n = 387$ cool neurons, 25 warm neurons; grey bars show comparison to tuned warm latency histogram in pIC). Vertical line represents the median. **d**, Three different example neurons (average of 5 trials) with transient (T) or sustained (S) responses to 10 °C cooling (top) or warming (bottom); histograms of duration index for cooling and warming stimuli. Arrowheads indicate the duration index of example neurons. **e**, Duration index of cool (top) and warm (bottom) responses plotted against the response latency. Green dashed line at duration index 0.5 highlights the separation between T and S neurons. **f**, Duration index of warm versus cool in broadly tuned neurons (n is as in **c** broadly tuned). **g**, Left: example traces of cool (T and S) and warm (S) neurons responding to ≈10-s stimulus at fast onset speed (left, about 130 °C s^{-1}) or at slower rate (right, about 1 °C s^{-1}) ($n = 5$ trials). Below, corresponding stimulus traces. Right: histograms of adaptation index for T and S neurons separated accordingly to 0.5 duration index as in **e** ($n = 241$ cool transient neurons, 147 cool sustained neurons, 98 warm sustained neurons, 6 mice, 11 sessions). Arrowheads indicate example neurons. **h**, Schematic model of how different channels of afferent input could drive cortical dynamics.

Source data

The spatial distribution of cortical neurons with respect to their functional properties is an important feature of cortical information processing[25,26,27]. To address whether pIC was organized thermotopically, we analysed the spatial distribution of thermally tuned neurons in pIC (Fig. 2f,g). Visual inspection showed a heterogeneous arrangement of thermally responsive neurons (Fig. 2f). In agreement, the distance between similarly tuned neurons (cool to cool; warm to warm) was not significantly different to the distance between differently tuned neurons (cool to warm; Fig. 2g). In a subset of experiments, we also used tactile stimulation and, in line with the differing spatial arrangements of touch and temperature, we observed a smaller fraction of thermally responsive pIC neurons (3%) than S1 neurons (12%) that responded to both thermal and tactile stimuli (Extended Data Fig.

3). Thus, thermal zones of pIC contain a heterogeneous, salt–pepper-like, arrangement of thermally responsive neurons.

Temporal dynamics of cortical thermal responses

The skin contains discrete spots that evoke cool or warm sensations[1,2,3]. Notably, in vivo recordings from sensory afferent neurons, which innervate thermal spots, have shown distinct response dynamics to thermal stimuli, with short-latency, transient cool response mediated by Aδ-fibres, and longer-latency, sustained firing in response to cool and warm stimuli by C-fibres[11,12,13,14]. We went on to investigate whether distinct temporal dynamics between cool and warm were also present in cortical responses (Fig. 3a). In remarkable similarity to the case for the periphery, a grand average of all significant cool and warm responding neurons (see Methods) in pIC showed a shorter-latency (about 80 ms) and more transient cool response compared to a longer-latency (about 320 ms) and more sustained warm response (Fig. 3b). The latency difference is present in both tuned and broadly tuned neurons (Fig. 3c), suggesting that thermal responses are driven by similar input. Moreover, the cool response latency in S1 is similar to that in pIC, whereas the sparse warm responses in S1 are substantially delayed and more variable compared to pIC warm responses (Fig. 3c and Extended Data Fig. 3).

We quantified the temporal dynamics of the thermal responses in pIC by computing a duration index that measures the change in response level at the end of the stimulation period compared to the initial peak value (Fig. 3d). The bimodality of the distribution of duration index for cool responses suggests a transient and a sustained response type (Fig. 3d). Similar results were observed for S1 cool responses (Extended Data Fig. 4). By contrast, pIC warm responses show a broad distribution of sustained responses (Fig. 3d). Plotting the duration index against the response latency showed that most cool transient neurons have a short-latency onset, whereas both cool and warm sustained neurons show delayed onsets (Fig. 3e). In broadly tuned neurons, we observed a similar distribution of cool and warm dynamics with cool transient and warm sustained responses in the same neurons (Fig. 3f), together highlighting that response dynamics are governed by sensory input rather than intrinsic cellular properties.

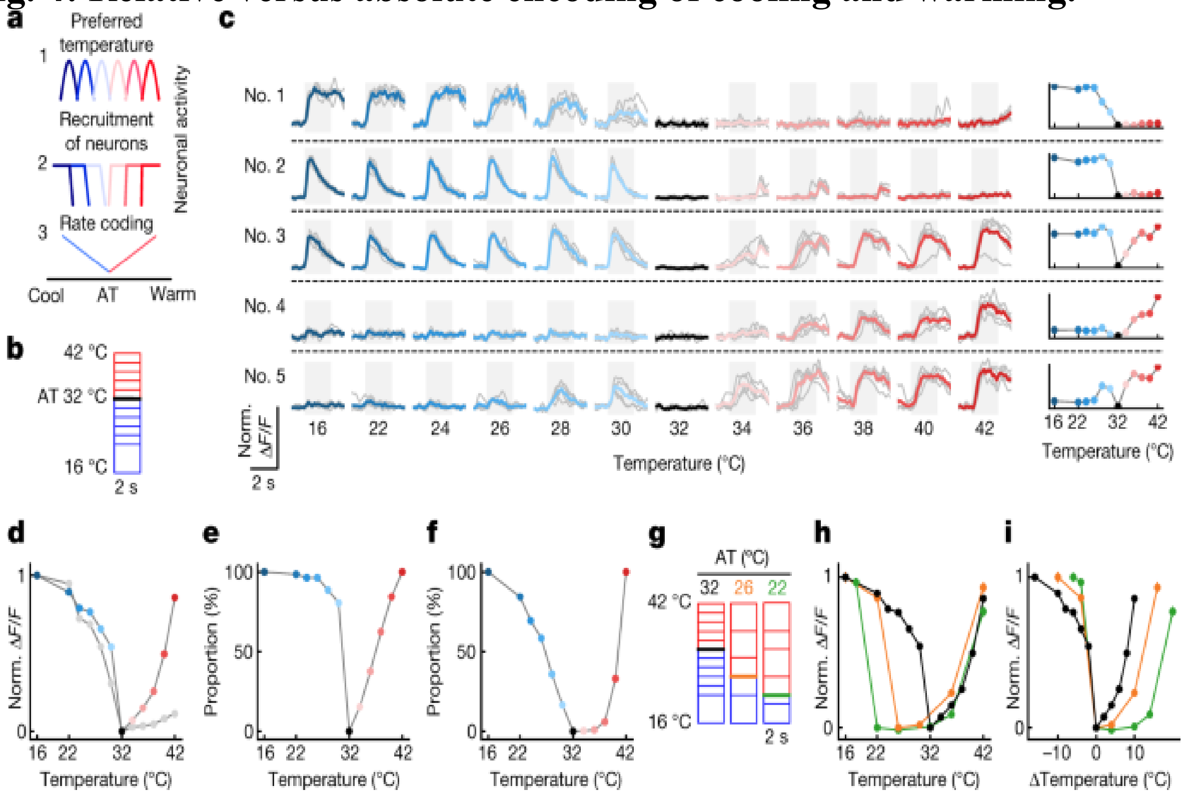
Fast-onset, transient neuronal responses are thought to be reliable indicators of stimulus change, whereas sustained responses are optimal for absolute stimulus level encoding. We tested this hypothesis by measuring the decrease in response amplitude when a thermal stimulus is presented with different onset rates (adaption index in Fig. 3g and Extended Data Fig. 4). We observed that cool transient neurons were not activated by stimuli with a slow onset whereas sustained neurons responded similarly irrespective of the stimulus onset speed. These features were observed in highly tuned and broadly tuned pIC neurons and also in S1 cool neurons (Extended Data Fig. 4). Overall, these data support a classic model of thermal encoding[2,3], whereby cool responses are driven by a combination of fast Aδ-fibre and slow C-fibre afferent input and warm by slower C-fibre input[11,12,13,14,18] (Fig. 3h).

Encoding of cooling and warming amplitude

The amplitude of a sensory stimulus can be encoded in different ways: by neurons with specific preferred amplitude values; the recruiting of more neurons as amplitude increases; graded rate coding (Fig. 4a). To address this in the thermal system, we measured single-cell responses in pIC to different amplitude cooling and warming stimuli (Fig. 4b) and show that responses to different amplitudes were reliable within a neuron, but diverse between neurons (Fig. 4c). Plotting the population response amplitude showed an asymmetric response profile with a steeper curve for cooling and more graded for warming (Fig. 4d). Similar data acquired in S1 showed a weak cellular representation of warm amplitudes (Fig. 4d, grey line). In pIC, most cool neurons respond to the smallest amplitude tested (2 °C) with only a minority recruited at larger amplitudes, whereas the number of activated warm neurons more gradually follows the increase of stimulus amplitude (Fig. 4e). The proportion of cool neurons reaching the maximal response amplitude is graded with stimulus amplitude, whereas warm neurons reach their maximum response amplitude only with large-amplitude stimuli (Fig. 4f). Overall, these analyses suggest that the encoding scheme for thermal amplitude is a combination of extra recruitment of neurons and graded changes of response amplitude. Among the neurons that responded significantly to warming, about 5% showed an unexpected response profile with thermometer-like properties, responding to small- but not large-

amplitude cooling. These ‘thermometer neurons’ showed graded reporting of absolute thermal value, independent of the stimulus direction from the AT (Fig. 4c, neuron no. 5, and Extended Data Fig. 5).

Fig. 4: Relative versus absolute encoding of cooling and warming.



a, Schematic showing three possible cortical thermal coding schemes. Line colours indicate neuronal activity in response to cooling (blue) and warming (red). **b**, Schematic of stimulus protocol from AT of 32 °C. **c**, Left: traces of example pIC neurons responding to cooling and warming from AT 32 °C; grey lines show individual trials (5 trials per temperature), and coloured lines show mean response. Right: peak response amplitude plotted as a function of the thermal stimulus for the example neurons on the left. **d**, Summary of response amplitude plotted as a function of the thermal stimulus for entire population of neurons ($n = 746$ neurons, 16 sessions, 7 mice). Coloured filled circles (pIC) and grey filled circles (S1) show mean ± s.e.m. at AT 32 °C. **e**, Same data as in **d**, but showing the proportion of recruited pIC neurons (>20% of response amplitude). **f**, Same data as in **d**, but showing the proportion of pIC neurons reaching maximum response (>80% of response amplitude). **g**, Schematic of stimulus protocols used to test impact of AT on thermal encoding. **h**, Graphs as in **d**, but for pIC

neurons at all ATs studied (AT 32 °C, *n* is same as **d**; AT 26 °C, $n = 401$ neurons, 9 sessions, 5 mice; AT 22 °C, $n = 448$ neurons, 10 sessions, 6 mice). **i**, Same data as **h**, but plotted against change in stimulus amplitude.

Source data

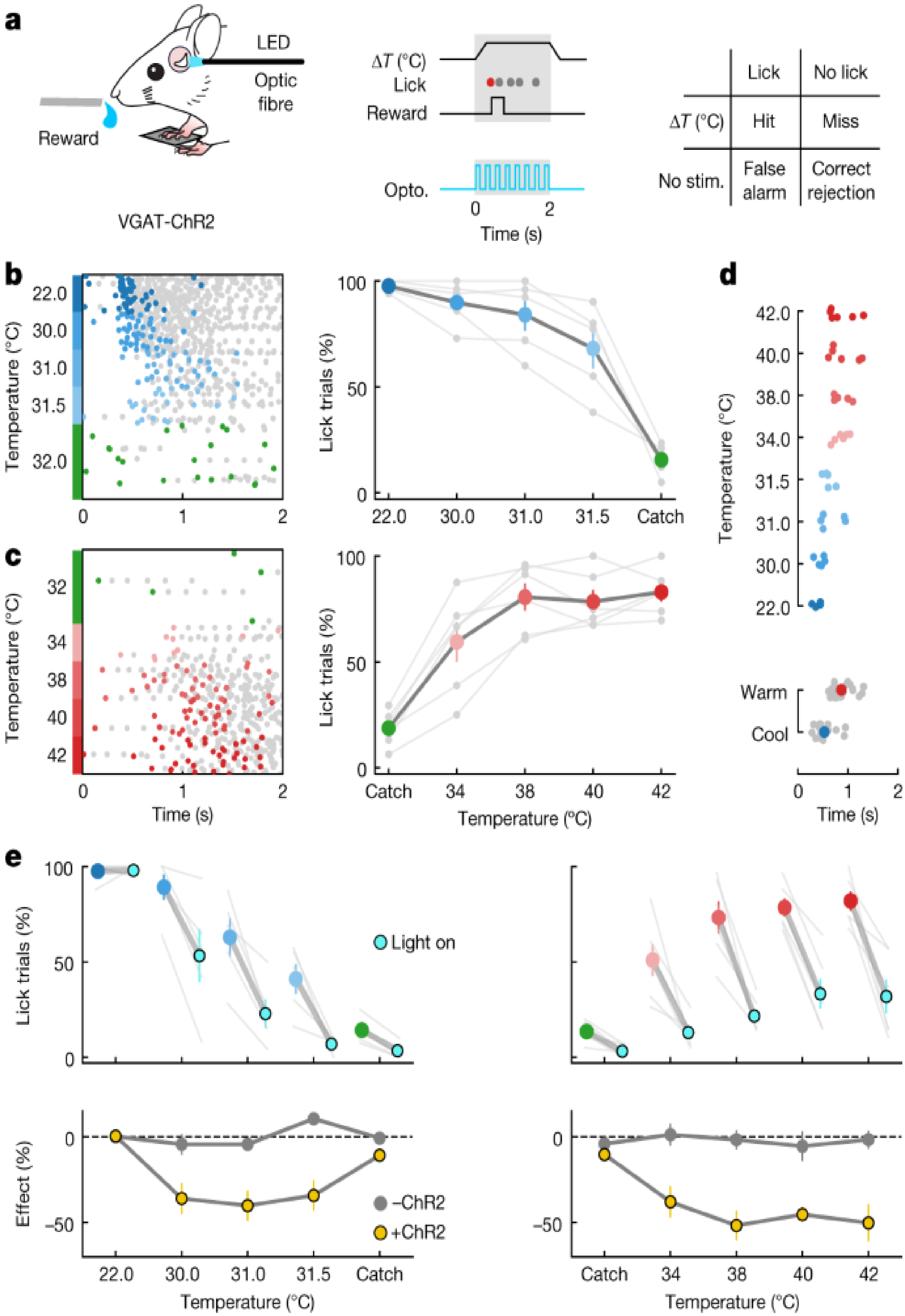
The thermal responses of primary sensory afferent neurons and spinal cord neurons are influenced by the skin AT[9,10,11,12,13,14,15,16,17], but whether this is reflected by a change in cortical response amplitude is not known. To investigate this, we compared pIC cellular responses to stimuli of the same amplitude but from ATs of 22 and 26 °C (Fig. 4g). At lower ATs, warming stimulus amplitude had to be substantially increased to observe response amplitudes comparable to those seen at AT 32 °C, and responses were often observed only once the stimulus target temperature reached 34 °C. By contrast, reliable responses to cooling were observed for all target temperatures and ATs tested (Fig. 4h). This difference was highlighted by plotting the response amplitude against the relative stimulus amplitude, which showed a shift in the warm response curve, but not in the cool (Fig. 4i, consistent conclusions were also observed in widefield imaging data in Extended Data Fig. 6). Together these data indicate that cooling and warming amplitude is encoded in fundamentally different ways in the cortex, with warming dependent on the absolute temperature and cooling reflecting the magnitude of temperature change, consistent with findings in primary sensory afferent neurons and spinal cord neurons[9,10,11,12,13,14,15,16,17].

Thermal perception is mediated by pIC

More anterior parts of insular cortex are thought to play a role in cognitive and motivational control of behaviour[28,29], whereas a central region is involved in taste perception[30]. Human lesions, microstimulation and imaging studies have suggested a role for insular cortex in thermal perception[5,20,31,32], but reversible manipulations have not been carried out and the link between thermal sensitivity and perception is unclear. To address this, we carried out optogenetic manipulations during a go/no-go thermal detection task (Fig. 5a). Thermal stimuli were delivered at random times from AT 32 °C. VGAT-ChR2 mice, with channelrhodopsin-2 (ChR2) constitutively expressed in GABAergic inhibitory interneurons, were

implanted with an optical window over pIC and provided with water rewards on licking a water spout during the onset ramp or plateau phase of the stimulus (Fig. 5a and Extended Data Fig. 7). Owing to differences in sensitivity to cooling and warming in pIC (Fig. 4), we chose 10 °C, 2 °C, 1 °C and 0.5 °C cooling stimuli, whereas for warming we used 10 °C, 8 °C, 6 °C and 4 °C. Mice were able to report stimuli at all amplitudes, highlighting their acute thermal perceptual ability (Fig. 5b,c). As for the functional responses (Figs. 1 and 3), mice reported warming at longer latencies than for cooling. We went on to inhibit the activity of pIC during the stimulus phase in the same mice (Fig. 5a,e) using light pulses (12.5 mW, 20 Hz, 2 s) through a 200-µm fibre optic positioned normal to the thermal region of pIC (Fig. 5a,e). In cooling-trained mice, optogenetic inhibition of pIC suppressed the hit rates to 2 °C, 1 °C and 0.5 °C stimuli but not to 10 °C (Fig. 5e), whereas, in warming-trained mice, optogenetic inhibition suppressed the hit rates to all amplitudes (Fig. 5e). Optical stimulation efficiently inhibited neuronal responses driven by thermal stimuli in pIC (Extended Data Fig. 7). Moreover, repeating the same optical stimulation paradigm but in Thy1-GCaMP6s mice not expressing ChR2 showed no effect on thermal perception (bottom panels Fig. 5e and Extended Data Fig. 7), confirming that the light alone does not alter the behaviour of the mice in our task.

Fig. 5: Thermal perception is mediated by pIC.


a
LED
Optic fibre
Reward
VGAT-ChR2
ΔT (°C)
Lick
Reward
Opto.
0
2
Time (s)
Lick
No lick
ΔT (°C)
Hit
Miss
No stim.
False alarm
Correct rejection
b
Temperature (°C)
22.0
30.0
31.0
31.5
32.0
Lick trials (%)
Catch
c
Time (s)
Temperature (°C)
d
Warm
Cool
e
Light on
Effect (%)
−ChR2
+ChR2

a, Left: schematic of thermal detection task and placement of optic fibre. LED, light-emitting diode. Middle: example trial structure showing timing of reward window (grey) and the timing of the optical stimulus during trials with optogenetic (opto.) manipulation (blue). Filled circles show licks; first rewarded lick is coloured. Right: response categories of task. stim., stimulation. **b**, Left: raster plot of licks in all trials ($n = 300$) from an example mouse trained for cooling sorted by stimulus amplitude; green filled circles show false alarms. Right: summary of behavioural performance showing proportion of trials with at least one lick in reward window ($n = 5$ mice). Grey lines show individual mice; coloured filled circles show mean ± s.e.m. **c**, Same as in **b**, but for warming (raster $n = 239$ trials, graph $n = 6$ mice). **d**, Top: latency of first lick is longer in response to warming than to cooling at different amplitudes. Filled circles show data from individual mice in **b**,**c**. Bottom: data from all amplitudes, with filled coloured circles showing mean ± s.e.m. **e**, Top left: proportion of trials with licks in VGAT-ChR2 mice ($n = 5$ mice) during optical stimulation (cyan) trials versus without optical stimulation (blue) for different cool stimulus amplitudes. Thin grey lines show individual mice; coloured filled circles joined by thick grey lines show mean ± s.e.m. Bottom left: the effect of light stimulus (change in percentage of trials with licks, mean ± s.e.m.) in VGAT-ChR2 mice (yellow), and in mice not expressing ChR2 (grey, see Extended Data Fig. 7). Right, same as at left, but for warm stimulation ($n = 6$ mice).

Source data

We next tested whether the differential impact of optogenetic inhibition of warming and cooling perception resulted from the additional representation of cool in S1. We found that the impact of optogenetic inhibition of S1 is weaker compared to that of pIC (Fig. 5e and Extended Data Fig. 7), supporting the hypothesis that pIC is the central representation of temperature and could be explained by differences in cortical area size, local cellular architecture and projection targets. As expected, simultaneous optogenetic inhibition of S1 and pIC using fibre optics positioned directly over pIC and S1 had a more pronounced effect on warm and cool perception than inhibiting S1 alone (Extended Data Fig. 7), but was similar to the impact of inhibiting pIC alone (Fig. 5e). Notably, even during dual inhibition, mice were still able to detect 10 °C cooling. This could result from the fast onset of cool responses providing a more robust encoding

strategy or perhaps reflect the recruitment of different upstream circuits. Future experiments should be designed to address these different hypotheses.

Optogenetic inhibition of pIC in VGAT-ChR2 mice during spontaneous licking of free water rewards and during an acoustic detection task did not alter lick rates (Extended Data Fig. 8). Moreover, optogenetic inhibition of pIC in VGAT-ChR2 mice trained to discriminate between an acoustic stimulus (go; rewarded) versus an acoustic stimulus presented simultaneously with a thermal stimulus (no-go; not rewarded) induced a selective increase of licking during the no-go trials, showing that pIC manipulation has an impact on perception rather than causing a general decrease of licking (Extended Data Fig. 8). Taken together, our behavioural data show that pIC plays a profound role in thermal perception.

Discussion

Here we identify a cortical region required for non-painful thermal perception and support the hypothesis that pIC houses a thermal cortex. Our data show that humans, monkeys and rodents therefore share not only similar perceptual abilities[18,24,33], but also a cortical area (pIC) specialized in temperature processing, putting the thermosensory system of mammals closer than previously thought[5,34]. Moreover, we observed distinct encoding features of cool and warm in the cortex. The representation of warm has delayed and uniform dynamics that encodes absolute stimulus amplitude, compared to the mixed temporal dynamics for cool that drive a relative encoding of stimulus amplitude. Similar features have been reported in primate and human primary thermal afferent neurons[11,12,13,14] and closely resemble thermal response features in *Drosophila* central and peripheral neurons[35,36], together highlighting the conserved nature of thermal encoding across the animal kingdom.

We find that cool and warm are represented in the cortex in a labelled-line-like fashion resembling sensory afferent encoding proprieties (Fig. 1a, top)[9,10,11,12,13,14,15,16,17]. Nevertheless, cortical neurons can also show response profiles not observed in the periphery, such as the encoding of temperature by thermometer neurons, raising the possibility of complex

thermal features being generated in the cortex (Fig. 4c and Extended Data Fig. 5). The presence of distinct features of cool and warm responses both in broadly and tightly tuned pIC neurons shows that functionally segregated streams of afferent information can merge in cortical neurons. This is interesting in light of recent data suggesting interactions between warm and cool afferent pathways[18] and findings from the tactile pathway that suggest substantial subcortical transformation of tactile information before reaching S1 (ref. [37]). Our findings will allow future research studies to examine the separation and integration of different channels of sensory afferent input on the cortical encoding of temperature.

Touch and cool response areas overlap in S1 (ref. [7]), whereas they seem more separate in pIC, hinting at differences both in the sensory input and functional output of these regions. Objects are normally cooler than skin temperature and thermal conductivity is an important component of object identification[38]. One hypothesis is that S1 integrates cool with touch during haptic sensing, whereas pIC encodes the thermal identity (warm versus cool) and level forming an independent and complete representation of thermosensation. The discovery of an optically accessible cortical representation of temperature provides a platform to address the neural mechanisms of non-painful thermal perception, as well as pain evoked during noxious stimulation, allodynia and thermal illusions.

Methods

Mice

All experimental procedures were carried out in accordance with the State of Berlin Animal Welfare requirements and European animal welfare law. Male and female mice older than 2 months were maintained on a 12:12 h light–dark cycle with experiments carried out during the light phase of the cycle. Mice were housed in groups at 22 ± 2 °C temperature and 55 ± 10% humidity with ad libitum access to food and water unless stated otherwise. Thy1-GCaMP6s (C57BL/6J-Tg(Thy1-GCaMP6s)GP4.3Dkim/J) mice were used for calcium imaging and for behavioural experiments (The Jackson Laboratory, stock no. 024275)[39,40]. VGAT-ChR2 (B6.Cg-Tg(Slc32a1-

COP4*H134R/EYFP)8Gfng/J) mice were used for behavioural experiments (The Jackson Laboratory, stock no. 014548)[41]. For awake experiments, mice were gradually habituated to head and paw fixation and tilted for optical access to pIC or S1.

Surgery

Mice were deeply anaesthetized (ketamine 120 mg kg^{-1} and xylazine 10 mg kg^{-1}) and injected subcutaneously with dexamethasone sodium phosphate (2 mg kg^{-1}) to prevent cerebral oedema, as well as metamizol (200 mg kg^{-1}). Anaesthetized mice were fixed with a nose-clamp, eye gel (Vidisic, Bausch + Lomb) was applied to both eyes and body temperature was maintained at 37 °C with a heating pad and rectal probe. After surgery, mice received a subcutaneous injection of warm sterile saline solution and were placed on a warm blanket until they awoke from anaesthesia. To avoid post-operative pain, metamizole was dissolved in the drinking water for 3 days post-surgery.

Implantation of cranial windows

To implant a window over pIC, mice were anaesthetized as above, and then the parietal and temporal bones of the left hemisphere were exposed and the left temporalis muscle was carefully separated from the temporal bone and partially removed. An approximately 3 × 3 mm craniotomy was carried out over pIC as identified by local anatomical landmarks (rhinal vein, middle cerebral artery, zygomatic bone). To implant a window over primary somatosensory forepaw cortex (S1) a 3-mm-diameter craniotomy was carried out over the left hemisphere S1 identified by the anatomical location respect to bregma (1.5–2.5 mm lateral and 0.5–1 mm anterior). To implant windows over pIC and S1 in the same mouse, we carried out two 3-mm craniotomies. The cortical surfaces were then rinsed with Ringer’s solution and two glass coverslips were placed on the surface of the cortex. The lower glass (diameter 3 mm or 3 × 3 mm) and the upper glass (diameter 4 or 5 mm) were glued together with optical adhesive (NOA 61, Norland Products). Dental cement and cyanoacrylate glue were used to attach a metal head post and the upper coverslip to the skull. Following pIC surgery, the skin lying over the temporal muscle was sutured.

Clear skull preparation

To gain simultaneous optical access to S1 and pIC, we used a through-skull preparation (Fig. 1f,g). Under deep isoflurane anaesthesia, the left temporalis muscle was partially removed and the dorsal surface of the skull was cleared of skin and periosteum. Finally, the exposed skull was sealed with cyanoacrylate glue.

Cortical pharmacology

For pharmacological inactivation, ≈500-μm-diameter craniotomies were drilled over cool (32–22 °C)-responsive areas in S1 and pIC in mice anaesthetized with isoflurane (1.5–2% in O_2). Both regions were identified functionally using the widefield calcium imaging response to thermal stimuli. Following the craniotomy, the dura was covered with transparent silicone gel (3-4680, Dow Corning). A 300 nl volume of muscimol or Ringer's solution was injected at a rate of 100 nl min^{-1} 300–500 μm below the pial surface using a pulled glass pipette and a hydraulic injection system (MO-10, Narashige). Muscimol (Abcam, ab120094) was dissolved in Ringer's solution to a concentration of 5 mM. Imaging sessions were carried out >10 min following the end of Ringer's solution or muscimol injection (Fig. 1f,g).

Sensory stimulation

Thermal

For widefield imaging, behavioural and electrophysiological experiments, thermal stimuli were delivered by an 8 × 8 mm Peltier element regulated by a feedback-controlled stimulator (Yale Medical School or a custom-made device, ESYS GmbH Berlin). Different thermal stimuli (2-s duration with an onset ramp speed of 20 °C s^{-1}) were interleaved randomly and delivered every 30 or 60 s. During paw stimulation, we took care to place the forepaw or hindpaw pad glabrous skin into direct contact with the centre of the Peltier element. During face stimulation, the Peltier surface was positioned on the orofacial region of the snout. For single-cell two-photon experiments,

thermal stimuli were delivered using a Peltier-based thermal stimulator (QST-lab). Thermal stimuli of different durations (2, 5, 10 s) or with different onset ramp speeds (approximately 130, 10, 3.3, 2, 1 °C s^{-1}) were interleaved randomly and delivered every 30 or 60 s.

Acoustic

Acoustic stimuli of 8 kHz, ≈65 dB sound pressure level and 1-s duration were delivered every 60 s using a loud speaker (Visaton) positioned 10 cm from the contralateral ear. The frequency of 8 kHz was chosen because it is well represented in the pIC auditory field[23]. For behavioural training, a 14 - kHz tone (≈65 dB, 1-s duration) was chosen as it is weakly represented in the pIC auditory field[23].

Vibrotactile

Vibrotactile stimuli (100 Hz) were generated by a piezoelectric actuator (PL127, PI) equipped with a 5-cm glass rod bent at the tip. The contralateral forepaw or hindpaw was tethered to a rigid support with a hole (diameter 5 mm) to allow the bent tip of the glass rod to stimulate the centre of the palm of the paw. In some experiments vibrotactile stimuli were generated using a small motor (Pololu shaftless vibration motor 8 × 3.4 mm) positioned on top of the paw tethered to the Peltier. Vibrotactile stimuli had a 0.5-s duration and were delivered every 30–60 s.

Widefield imaging

Setup

Imaging was carried out with a Leica MZ10F stereomicroscope. Blue excitation light (470 nm) was emitted from a LED (pE-300, CoolLED) and bandpass filtered (470/40 nm). Emission light was bandpass filtered (525/50 nm) before recording with a scientific complementary metal–oxide–semiconductor camera (ORCA-Flash4.O LT, Hamamatsu). Images were acquired at a rate of 20 Hz and 35 ms exposure time. Frame size was 1,024 × 1,024 pixels (2 × 2 binning); for cranial window preparation the field of

view was about 2.7 × 2.7 mm and for the cleared skull preparation the field of view was about 6.7 × 6.7 mm. Acquisition trials lasted 10–15 s and had an inter-trial-interval time of 30–60 s. During pIC imaging, animals were tilted to allow optical access. As a result of physical constraints, pIC and S1 widefield imaging was carried out in different animals and sessions, except during the cortical pharmacological inactivation experiments (Fig. 1f,g). In this case, we used a 'clear skull' preparation and imaged pIC and S1 simultaneously. The setup was always adjusted to allow the paw to be placed in a similar position to all other experiments. Data acquisition was controlled by custom-written code (Python).

Widefield image analysis

Videos were motion corrected, and areas in the recordings outside the brain were masked. We calculated the relative change in fluorescence as $\Delta F/F = (F(t) - F_0)/F_0$, in which F_0 is the 15th percentile of the activity in the trial. The average activity of the last 500 ms before stimulus was subtracted from $\Delta F/F$. The activity of a given region was quantified by averaging the fluorescence in a region of interest (ROI; diameter 15 pixels) before calculating $\Delta F/F$. When placing ROIs, we avoided visible blood vessels, but otherwise positioned the ROI at the area with the peak stimulus-driven response. The ROI positions were kept the same over days and recording sessions by comparing the position relative to the blood vessels. Auditory stimuli activated several fields in the auditory cortex and one area in pIC (insular auditory field).

Somatotopic maps

Somatotopic maps of sensory input to pIC (Fig. 1) and S1 (Extended Data Fig. 2) are grand averages across mice and trials. Not all stimuli were delivered to all mice. Using the ROI of the thermal forepaw response as a reference location, $\Delta F/F$ videos for all trials were aligned by translation. The grand averages were smoothed with Gaussian filter ($\sigma = 20$ pixels) and thresholded to show only peak activity (>85% of the maximum activity for thermal and >90% for tactile stimulation). For pIC, because of the simultaneous responses in the auditory insular field and auditory cortex to sound stimuli, responses were analysed independently. We carried out this

analysis only on data with grand average peak response amplitude >3% $\Delta F/F$.

Response onset estimation

We defined the onset of the widefield response as the peak of the second derivative of the widefield signal in a time window from the start of stimulus to the peak of the first derivative. If negative values of the first derivative appeared after stimulus start, the time window was from the last of such negative values. We carried out this analysis only on data with a peak response amplitude >3% $\Delta F/F$.

Two-photon imaging

Setup

Fluorescence signals were recorded using a two-photon microscope (ThorLabs Bergamo II, 12 kHz scanner) with a Nikon 16× water-immersion objective (NA 0.8) giving a field of view of 540 × 540 μm. The scope was operated using ThorImageLS software (v4.0.2019.8191, Thorlabs). In most experiments the microscope was rotated by 45° from vertical. Owing to physical constraints, pIC and S1 two-photon imaging was carried out in different animals and sessions. The excitation laser (InSight DeepSee, Spectra-Physics) was tuned to 940 nm, and the power never exceeded 100 mW. Emitted photons were bandpass filtered 525/50 (green) onto a GaAsP photomultiplier tube. Multi-plane (512 × 512 pixels) acquisition was controlled by a fast piezoelectric objective scanner, with planes spaced 45 μm apart in depth. Seven planes were acquired sequentially, and the scanning of the entire stack was repeated at about 5 Hz. Beam turnarounds at the edges of the image were blanked. Acquisition trials lasted 26 s and had an inter-trial interval of 30–60 s. Stimuli generation and hardware synchronization were carried out on a computer with a National Instrument card running custom-written Python code.

Two-photon analysis

Motion correction of data, identification of putative neurons and calculation of $\Delta F/F$ was carried out using the Suite2p package (v0.9.3) in Python[42]. Identified neurons were manually verified by visual examination of the traces and the spatial footprints of the neurons. A Savitzky–Golay filter was applied to traces presented in figures for visual purpose alone.

Criteria for responsive neurons

Neurons with a significant increase in activity to 10 °C cooling and warming stimuli were included for further analysis. Cool and warm neurons were defined as those significantly responding to either 10 °C cooling or warming stimuli. Broadly tuned neurons had significant response to both cooling and warming stimuli. To identify an increase in response amplitude, the activity in a time window before stimulus onset was compared with the activity in a time window during stimulus. A neuron was identified as being responsive if the change in activity was significantly larger (estimated by bootstrapping) than two times the noise level of the neuron measured by the interquartile range of activity before the stimuli.

The thermal bias index (Fig. 2d,e) for a thermally responding neuron is the normalized difference between its cool and warm responses, (warm − cool)/(warm + cool), in which 'warm' is the average peak response to 32–42 °C and 'cool' is the average peak response to 32–22 °C. Grand average (Fig. 3b,c) calculated from all significantly responding neurons shown in Fig. 2 and includes 470 neurons with significant responses to cooling and 401 to warming. Of these, 125 responded to both. The duration index (Fig. 3d) measures the change in the fluorescence level between the initial peak value (f_{init}) and the end (f_{end}) of a 2-s thermal stimulus, $(f_{\mathrm{init}} - f_{\mathrm{end}})/f_{\mathrm{init}}$. The adaptation index (Fig. 3g) was calculated as the difference in the maximum fluorescence level during fast thermal stimulation (about 130 °C s^{-1}; f_{fast}) and during slow thermal stimulation (≈1 °C s^{-1}; f_{slow}), $(f_{\mathrm{fast}} - f_{\mathrm{slow}})/f_{\mathrm{fast}}$.

Behavioural experiments

Setup

Head-fixed mice (VGAT-ChR2 and Thy1-GCaMP6s) were implanted with a glass window over pIC and trained on a go/no-go stimulus detection paradigm. Mice were rewarded with a water drop (about 2 µl) if they reported a randomly timed thermal stimulus with at least one lick of a water spout (capacitance sensor) within a window of opportunity (from the start of thermal stimulus to start of offset ramp). Correct rejections were not rewarded and incorrect responses were not punished, although premature licking in the 5 s before the stimulus onset would postpone the next trial by 5 s. Mice were water restricted and their weight was monitored daily. For behavioural training, a 200-µm-diameter, 0.22-NA optic fibre (Thorlabs) was coupled to an LED (470 Plexon LED source) and placed on the coverslip surface orthogonal to the thermal region of the pIC or S1 using the blood vessel pattern. In a subset of mice, cranial windows were implanted above both S1 and pIC, and LEDs were placed above both areas (Extended Data Fig. 7). Control of behavioural training and data collection was carried out using either custom-written Labview software (16.0f5, National Instruments, USA) or the Bpod system (1.8.2, Sanworks, USA).

Thermal perception task

Thermal perception (Fig. 5) training involved several stages: (1) Free access to water rewards from the water spout; (2) automatic water rewards paired to the presentation of 10 °C, 2-s cooling and warming stimuli from AT 32 °C; (3) training to report a 10 °C cooling and warming stimuli by licking for a water reward within a 2 s window from the start to the end of the plateau phase of the stimulus; thermal stimuli trials and catch trials were presented with a ratio of 1:1 with an inter-stimulus-interval of 15–20 s; (4) once a hit rate >70% and false alarm rate <30% had been reached, 4 randomized stimulus amplitudes were used from AT 32 °C (cooling, 22 °C, 30 °C, 31 °C and 31.5 °C; warming 42 °C, 40 °C, 38 °C and 34 °C) with a 4:2 ratio of stimulus trials to catch trials; (5) once a hit rate of >70% and false alarm rate <30% had been achieved for at least 1 amplitude, in the next session we presented the same 4 thermal stimuli with 1 catch trial with LED on or off (on/off trials) with a ratio of 2:1 (randomized). The LED was on for the entire duration of the stimulus and delivered at 20 Hz, 50% duty cycle with a power of 12.5 mW (measured at tip of fibre). Before the start of stage (4) or

(5) sessions, mice were exposed to a brief session of the stage (3) protocol to quench initial thirst.

Thermal discrimination task

The thermal discrimination task training (Extended Data Fig. 8) involved several stages: (1) free access to water rewards from the water spout; (2) automatic water rewards paired to the presentation of 14-kHz, ≈65-dB, 0.5-s acoustic stimulus; (3) training to report 14-kHz stimulus by licking for a water reward within a 1.4-s window from the start of the stimulus (go trial); acoustic stimuli (14 kHz, about 65 dB) delivered 0.6 s after the beginning of a 2-s thermal stimulus (either cooling or warming) (no-go trial) and catch trials (no stimuli) were not rewarded; go, no-go and catch trials were presented with a ratio of 1:1:1 with an inter-stimulus-interval of 15–20 s; training was continued until go trial hit rate >70%, no-go false alarm rate around 50% and false alarm catch rate <30%; (4) once a mouse learned to discriminate between acoustic stimulus alone (go) and acoustic stimulus presented together with thermal stimuli (no-go), in the next session we presented the same stimuli interleaved at a ratio of 2:1 into trials with the LED off (off) and trials with the LED on (on). The LED was on for the entire duration of the thermal stimulus and delivered at 20 Hz, 50% duty cycle with a power of 12.5 mW (measured at tip of fibre). Before the start of stage (4) sessions, mice were exposed to a brief session of the stage (3) protocol to quench initial thirst.

Acoustic perception

The behaviour training to report acoustic stimuli (Extended Data Fig. 8) had the same structure as the thermal training above, but used a 14-kHz, ≈65-dB, 2-s acoustic stimulus.

Free licking

To monitor the impact of pIC optogenetic manipulation on licking behaviour, water restricted mice were allowed to freely lick from a lick spout with continuous rewards. Licking rates during the 2 s before the onset

of the light stimulus (Off) were compared with those during the 2 s after the onset of the light stimulus (On). Light stimuli were delivered for 5 s at 12.5 mW, 20 Hz and 50% duty cycle (Extended Data Fig. 8).

Datasets and analysis

Mice were randomly allocated to be trained first to report cooling or warming. In a second training round, they were trained to report the opposite stimulus. Data measuring the thermal perceptual ability of mice without optogenetic manipulation (Fig. 5b–d) were generated from the same mice used in optogenetic manipulations (Fig. 5e,f), the day before optogenetic testing. The effect of the optogenetic manipulation (bottom panels of Fig. 5e,f and Extended Data Fig. 7) was quantified by the mean change in the percentage of lick trials.

In vivo electrophysiology

Electrophysiological recordings for measuring the impact of optogenetic inhibition on thermal responses were carried out in VGAT-ChR2 mice implanted with a head post. Mice were habituated for head and paw fixation. On the day of the recording, the mouse was briefly anaesthetized (about 30 min, 1% isoflurane) and a craniotomy of about 2 mm diameter was carried out over the pIC. The dura was removed in the ventral part of the craniotomy and the brain was covered with Kwik-Cast silicon (World Precision Instruments). The animal was placed in its home cage to recover from anaesthesia for at least 2–3 h before head fixation in the recording setup.

We carried out 7 recordings in 4 mice using 32-channel silicon probes (Neuronexus) in either a linear or four-shank ‘Buzsaki32’ configuration. In some experiments, the tips of the probe were covered with DiI (Sigma). The probe was inserted orthogonal to the pIC with a micromanipulator (Luigs&Neumann) at 1–2 $\mu m\ s^{-1}$. Recordings were acquired at 30 kHz with a Neuralynx system (Cheetah). The extracellular recordings were spike sorted using Kilosort (version 2)[43]. The data were collected to evaluate the effect of ChR2 on neural activity, so we carried out limited manual sorting to exclude obvious noise but did not take any steps to exclude multi-unit

activity. Units with a mean activity during stimulus representation exceeding a threshold of 2 s.d. above the background were considered to be thermally responsive.

Two-second-long thermal stimuli of 10 °C cooling and warming were interleaved randomly and delivered every 30 s. The LED fibre (200-µm-diameter, 0.22-NA optic fibre, Thorlabs) was tilted slightly to accommodate the silicone probe. The LED was on for the entire duration of the thermal stimulus and delivered at 20 Hz, 50% duty cycle with a power of 12.5 mW (measured at tip of fibre).

Histology

To label the thermal cortical representation, animals were deeply anaesthetized by intraperitoneal injection of ketamine and xylazine. The centre of the response area was marked with a glass pipette painted with fluorescent dye (DiI, 5 mg ml^{-1}). Mice were then perfused transcardially using cold phosphate-buffered saline (PBS; 0.1 M, pH 7.4) and fixed with paraformaldehyde (PFA; 2% in 0.1 M PBS). The brain was removed and kept for 1–5 h in 2% PFA. After washing with PBS, the hemispheres were separated and cortices were removed and flattened between two glass slides separated by a spacer of 1–2 mm. Glass slides were weighed down for approximately 3–8 h at 4 °C in 2% PFA. After washing with PBS, 70–80-µm sections were cut on a Vibratome (Leica VT1000s). Sections were stained for cytochrome oxidase activity (2 mg cytochrome *c*, 6 mg diaminobenzidine). After the staining procedure, sections were mounted on glass slides with Mowiol mounting medium. Images were acquired with a Zeiss microscope (AX10) using a 5× objective and processed using Fiji (ImageJ, NIH, USA). Area borders were manually delineated following the contrast of the cytochrome oxidase stain and are comparable to previously reported data[44]. Coronal sections of 50 µm in thickness of PFA-fixed mouse brains were stained for 48 h with the following primary antibodies: mouse anti-Gad67 (catalogue number MAB5406; clone 1G10.2; Millipore; 1:800); chicken anti-GFP (catalogue number ab13970; Abcam; 1:250); mouse anti-NeuN (catalogue number MAB377; clone A60; Millipore; 1:100). The secondary antibodies Cy3 goat anti-mouse (A-21422; Invitrogen; 1:250) and A488 goat anti-chicken (A-11039; Invitrogen; 1:250) were incubated for a

few hours at room temperature. Cell counting was carried out manually in Fiji using the Cell Counter plug-in on epifluorescence images acquired using a Zeiss microscope (AX10) using a 10× objective.

Data analysis and statistics

No statistical methods were used to predetermine sample sizes, but our sample sizes were similar to those used in previous publications. Experimenters were not blind to trial order, but trial order was randomized during the experiment. All data analysis was carried out using Python. Uncertainty of means are reported with standard error of the mean. Bootstrapping was used for estimating central 95% confidence intervals using 5,000 resamples. Some histograms in Fig. 3 and Extended Data Fig. 4 do not show outlier data points. In Fig. 1f,g, we used two-sided paired *t*-test. For Fig. 1f: cool S1, $t = -3.78$, $P = 0.0323$; warm S1, $t = -0.17$, $P = 0.8767$; cool pIC, $t = 0.44$, $P = 0.6929$; warm pIC, $t = -0.45$, $P = 0.6839$. For Fig. 1g: cool S1, $t = 0.55$, $P = 0.6182$; warm S1, $t = -0.68$, $P = 0.5454$; cool pIC, $t = -3.61$, $P = 0.0366$; warm pIC, $t = -8.52$, $P = 0.0034$.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are available from the corresponding authors upon request. Source data are provided with this paper.

Code availability

Code is available from the corresponding authors upon request.

References

1. Blix, M. Experimentela bidrag till lösning af frågan om hudnervernas specifika energi. *Upsala Läkarefören. Förhandlin.* **18**, 87–102 (1882).

2. Filingeri, D. in *Comprehensive Physiology* (ed. Terjung, R.) 1429–1491 (Wiley, 2016).

3. Vriens, J., Nilius, B. & Voets, T. Peripheral thermosensation in mammals. *Nat. Rev. Neurosci.* **15**, 573–589 (2014).

4. Bokiniec, P., Zampieri, N., Lewin, G. R. & Poulet, J. F. A. The neural circuits of thermal perception. *Curr. Opin. Neurobiol.* **52**, 98–106 (2018).

5. Craig, A. D., Chen, K., Bandy, D. & Reiman, E. M. Thermosensory activation of insular cortex. *Nat. Neurosci.* **3**, 184–190 (2000).

6. Hellon, R. F., Misra, N. K. & Provins, K. A. Neurones in the somatosensory cortex of the rat responding to scrotal skin temperature changes. *J. Physiol.* **232**, 401–411 (1973).

7. Milenkovic, N. et al. A somatosensory circuit for cooling perception in mice. *Nat. Neurosci.* **17**, 1560–1566 (2014).

8. Tsuboi, Y. et al. Response properties of primary somatosensory cortical neurons responsive to cold stimulation of the facial skin and oral mucous membrane. *Brain Res.* **613**, 193–202 (1993).

9. Andrew, D. & Craig, A. D. Spinothalamic lamina I neurones selectively responsive to cutaneous warming in cats. *J. Physiol.* **537**, 489–495 (2001).

10. Craig, A. D., Krout, K. & Andrew, D. Quantitative response characteristics of thermoreceptive and nociceptive lamina I spinothalamic neurons in the cat. *J. Neurophysiol.* **86**, 1459–1480 (2001).

11. Darian-Smith, I., Johnson, K. O. & Dykes, R. ‘Cold’ fiber population innervating palmar and digital skin of the monkey: responses to cooling pulses. *J. Neurophysiol.* **36**, 325–346 (1973).

12. Darian-Smith, I. et al. Warm fibers innervating palmar and digital skin of the monkey: responses to thermal stimuli. *J. Neurophysiol.* **42**, 1297–1315 (1979).

13. Duclaux, R. & Kenshalo, D. R. Sr Response characteristics of cutaneous warm receptors in the monkey. *J. Neurophysiol.* **43**, 1–15 (1980).

14. Kenshalo, D. R. & Duclaux, R. Response characteristics of cutaneous cold receptors in the monkey. *J. Neurophysiol.* **40**, 319–332 (1977).

15. Ran, C., Hoon, M. A. & Chen, X. The coding of cutaneous temperature in the spinal cord. *Nat. Neurosci.* **19**, 1201–1209 (2016).

16. Wang, F. et al. Sensory afferents use different coding strategies for heat and cold. *Cell Rep.* **23**, 2001–2013 (2018).

17. Yarmolinsky, D. A. et al. Coding and plasticity in the mammalian thermosensory system. *Neuron* **92**, 1079–1092 (2016).

18. Paricio-Montesinos, R. et al. The sensory coding of warm perception. *Neuron* **106**, 830–841 (2020).

19. Beukema, P. et al. TrpM8-mediated somatosensation in mouse neocortex. *J. Comp. Neurol.* **526**, 1444–1456 (2018).

20. Penfield, W. & Faulk, M. E. The insula: further observations on its function. *Brain* **78**, 445–470 (1955).

21. Gogolla, N., Takesian, A. E., Feng, G., Fagiolini, M. & Hensch, T. K. Sensory integration in mouse insular cortex reflects GABA circuit maturation. *Neuron* **83**, 894–905 (2014).

22. Rodgers, K. M., Benison, A. M., Klein, A. & Barth, D. S. Auditory, somatosensory, and multisensory insular cortex in the rat. *Cereb. Cortex* **18**, 2941–2951 (2008).

23. Sawatari, H. et al. Identification and characterization of an insular auditory field in mice:insular auditory field in mice. *Eur. J. Neurosci.*

34, 1944–1952 (2011).

24. Stevens, J. C. & Choo, K. K. Temperature sensitivity of the body surface over the life span. *Somatos. Mot. Res.* **15**, 13–28 (1998).

25. Chen, K., Kogan, J. F. & Fontanini, A. Spatially distributed representation of taste quality in the gustatory insular cortex of behaving mice. *Curr. Biol.* **31**, 247–256 (2021).

26. Chen, X., Gabitto, M., Peng, Y., Ryba, N. J. P. & Zuker, C. S. A gustotopic map of taste qualities in the mammalian brain. *Science* **333**, 1262–1266 (2011).

27. Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).

28. Craig, A. D. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* **3**, 655–666 (2002).

29. Gogolla, N. The insular cortex. *Curr. Biol.* **27**, R580–R586 (2017).

30. Peng, Y. et al. Sweet and bitter taste in the brain of awake behaving animals. *Nature* **527**, 512–515 (2015).

31. Birklein, F., Rolke, R. & Müller-Forell, W. Isolated insular infarction eliminates contralateral cold, cold pain, and pinprick perception. *Neurology* **65**, 1381–1381 (2005).

32. Mazzola, L., Isnard, J. & Mauguière, F. Somatosensory and pain responses to stimulation of the second somatosensory area (SII) in humans. A comparison with SI and insular responses. *Cereb. Cortex* **16**, 960–968 (2006).

33. Rózsa, A. J., Molinari, H. H., Greenspan, J. D. & Kenshalo, D. R. The primate as a model for the human temperature-sensing system: 1. Adapting temperature and intensity of thermal stimuli. *Somatosens. Res.* **2**, 303–314 (1985).

34. Craig, A. D. A rat is not a monkey is not a human: comment on Mogil (*Nature Rev. Neurosci.* 10, 283–294 (2009)). *Nat. Rev. Neurosci.* **10**, 466–466 (2009).

35. Frank, D. D., Jouandet, G. C., Kearney, P. J., Macpherson, L. J. & Gallio, M. Temperature representation in the *Drosophila* brain. *Nature* **519**, 358–361 (2015).

36. Liu, W. W., Mazor, O. & Wilson, R. I. Thermosensory processing in the *Drosophila* brain. *Nature* **519**, 353–357 (2015).

37. Emanuel, A. J., Lehnert, B. P., Panzeri, S., Harvey, C. D. & Ginty, D. D. Cortical responses to touch reflect subcortical integration of LTMR signals. *Nature* **600**, 680–685 (2021).

38. Ho, H.-N. & Jones, L. A. Contribution of thermal cues to material discrimination and localization. *Percept. Psychophys.* **68**, 118–128 (2006).

39. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).

40. Dana, H. et al. Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo. *PLoS ONE* **9**, e108697 (2014).

41. Zhao, S. et al. Cell type–specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat. Methods* **8**, 745–752 (2011).

42. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at *bioRxiv* https://doi.org/10.1101/061507 (2016).

43. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris K. D. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. Preprint at *bioRxiv* https://doi.org/10.1101/061481 (2016).

44. Gămănuț, R. et al. The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. *Neuron* **97**, 698–715 (2018).

45. Makino, H. et al. Transformation of cortex-wide emergent properties during motor learning. *Neuron* **94**, 880–890 (2017).

46. Zhang, M., Kwon, S. E., Ben-Johny, M., O'Connor, D. H. & Issa, J. B. Spectral hallmark of auditory-tactile interactions in the mouse somatosensory cortex. *Commun. Biol.* **3**, 64 (2020).

47. Bokiniec, P., Whitmire, C. J., Leva, T. M. & Poulet, J. F. A. Brain-wide connectivity map of mouse thermosensory cortices. *Cereb. Cortex* https://doi.org/10.1093/cercor/bhac386 (2022).

48. Craig, A. D. in *Handbook of Clinical Neurology* Vol. 156 (ed. Romanovsky, A. A.) 317–338 (Elsevier, 2018).

49. Mazzola, L., Faillenot, I., Barral, F.-G., Mauguière, F. & Peyron, R. Spatial segregation of somato-sensory and pain activations in the human operculo-insular cortex. *NeuroImage* **60**, 409–418 (2012).

50. Chen, X.-J., Liu, Y.-H., Xu, N.-L. & Sun, Y.-G. Multiplexed representation of itch and mechanical and thermal sensation in the primary somatosensory cortex. *J. Neurosci.* **41**, 10330–10340 (2021).

Acknowledgements

This work was supported by the European Research Council (ERC-2015-CoG-682422, J.F.A.P.), the European Union (3x3Dimaging 323945, J.F.A.P.), the Deutsche Forschungsgemeinschaft (FOR 2143, J.F.A.P.; SFB 1315, J.F.A.P.), the Helmholtz Society (J.F.A.P.), the Centre National de la Recherche Scientifique (M.C.), the Independent Research Fund Denmark (M.V.) and the MDC-Weizmann Helmholtz International Research School iNAMES (G.G.). We thank M. Brecht, G. Lewin, J. Kremkow, S. Crochet and members of the laboratory of J.F.A.P. for constructive comments on an earlier version of the manuscript, S. Steinfelder for help with administrative and technical aspects, J. König for help with immunohistochemistry, B.

Eickholt for sharing NeuN antibody and A. Burkhalter for advice on the flattened cortex preparation and cortical parcellation.

Funding

Open access funding provided by Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC).

Author information

Author notes

1. These authors contributed equally: M. Vestergaard, M. Carta

Authors and Affiliations

1. Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

   M. Vestergaard, M. Carta, G. Güney & J. F. A. Poulet

2. Neuroscience Research Center, Charité-Universitätsmedizin Berlin, Berlin, Germany

   M. Vestergaard, M. Carta, G. Güney & J. F. A. Poulet

3. Univ. Bordeaux, CNRS, IINS, UMR 5297, Bordeaux, France

   M. Carta

4. Humboldt-Universität zu Berlin, Institut für Biologie, Berlin, Germany

   G. Güney

Contributions

M.V., M.C. and J.F.A.P. designed the study. M.V. M.C. and G.G. carried out experiments and analysed the data. M.V., M.C. and J.F.A.P. wrote the manuscript.

Corresponding authors

Correspondence to M. Vestergaard, M. Carta or J. F. A. Poulet.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Nadine Gogolla, Zachary Knight and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Functional and histological localization of pIC.

a, Functional and histological localization of the thermal zone of forepaw pIC in 2 example mice. Left, in vivo image of cortical surface of a Thy1-GCaMP6s mouse with superimposed, coloured areas showing the widefield

responses to cool (blue, 32-22 °C) and sound (green, 8 kHz). Glass pipette painted with DiI (5 mg/ml) was inserted in the center of the cool response area (black cross). Right, cytochrome oxidase staining of a flattened cortex preparation from the same mouse on left with site of DiI injection (magenta). Areal borders were manually delineated based on CO intensity reaction, parcellation followed[44]. **b**, Second example mouse as in **a**. **c**, Left, in vivo image of cortical surface of a Thy1-GCaMP6s mouse with superimposed, coloured areas showing the widefield responses to cool (blue, 32-22 °C) and sound (green, 8 kHz). Center, coronal section of the same mouse on the left containing the site of DiI injection (magenta). GCaMP6s signal was amplified with immunohistochemistry with anti-GFP antibody. Right, zoom of the injection site. Example mouse is the same presented in Fig. 2c. Similar results were found in other two mice. **d**, Top, example of epifluorescence image of pIC from a Thy1-GCaMP6s mouse stained for GCaMP6s (labelled with anti-GFP antibody) and for the general neuronal marker NeuN. Bottom, cell density of NeuN- and GCaMP-positive neurons. Right, quantification of the percentage of NeuN-positive neurons that also expressing GCaMP6s in the imaged section across layers (total) and in the imaging layer (2p, 100–370 µm) (n = 3 mice, mean). Values are in line with previous data[40]. Filled circles show data from different mice. **e**, Same as **d**, but with staining for GCaMP6s and with antibody for GABAergic neurons (GAD67). Right, percentage of GAD67-positive neurons among GCaMP6s-expressing neurons (n = 4 mice, mean). Values are in line with previous data[45]. **f**, Cartoon schematic representing the position of the acoustic and forepaw cool responsive fields relative to major blood vessels. IAF, insular auditory field; AC, auditory cortex. Different terminology has been used for this region in previous studies, including pIC, parietal ventral area (PV) or S2/IC, highlighting the difficulties in discerning two neighboring cortical regions. In agreement with studies that have used functional mapping of sensory responses[22,23,29,46,47] and because pIC is a region associated with thermal processing in human studies[2,4,5,31,48,49] here we use the term pIC.

Source data

Extended Data Fig. 2 Somatotopic representation of temperature in S1.

a, Cartoon schematic showing the location of the forepaw cool responsive area (blue) relative to the major blood vessels. **b**, From left to right, cartoon schematic showing imaging setup; somatotopic map of responses locations to thermal (10 °C cool) and tactile. Coloured area indicates peak population response averaged across mice (see methods, *n* = 12 thermal forepaw, 5 thermal hindpaw, 4 thermal face, 12 touch forepaw, 5 touch hindpaw). Data from individual mice are aligned to peak activity of the thermal forepaw response. **c**, Widefield responses to thermal stimulation of different body parts from the same dataset as in **b**. Grey lines show mean responses from individual mice (*n* = same as **b**), coloured lines show population mean, grey area indicates time from start of stimulus to end of plateau phase. **d**, Grey filled circles show response latencies from individual mice, coloured filled circles show mean ± s.e.m., same data as in **c**.

Source data

Extended Data Fig. 3 Cellular encoding of temperature in S1.

a, Cartoon schematic showing two-photon imaging of S1. Imaging started at 100 µm from the pial surface of a Thy1-GCaMP6s mouse, 7 optical sections were acquired with intervals of 45 µm. **b**, Top, example in vivo two photon image of GCaMP6 expressing neurons in S1. Bottom, responses of three example S1 neurons during 10 °C cooling (blue) or 10 °C warming (red) stimuli. Here and in all figures, grey lines show single trial responses, coloured lines show average, grey shaded box indicates time from start of stimulus to end of plateau phase. Below, corresponding thermal stimulus traces. **c**, Single neuron S1 calcium responses to cooling (left) and warming (right) stimuli. Each line represents a single neuron, responses are normalized to the peak and sorted based on the thermal bias index with white lines showing the onset and the end of the plateau phase of thermal stimuli (n = 411 neurons, 4 mice, 9 sessions). **d**, Bar graph showing the percentage of pIC and S1 neurons responding to thermal only and also to tactile stimulation (pIC, 506 neurons thermal only, 18 neurons thermal and tactile; S1, 345 neurons thermal only, 49 neurons thermal and tactile). The overall number and proportion of thermal and tactile neurons in S1 presented here differs from a previous report[50]. This might depend on the body part studied, stimulus design and recording conditions.

Source data

Extended Data Fig. 4 Dynamics of thermal responses to thermal stimuli with different stimulus onset speeds.

a, Responses of example transient and sustained cool (T and S) and sustained warm (S) neurons responding to stimuli with different onset speeds of approximately 130, 10, 3, 2, 1 °Cs^{-1}. Example neurons are the same as presented in main Fig. 3g. Grey lines show individual trials, coloured lines show mean response. Dashed lines show mean response of the corresponding neuronal population. **b**, Responses of example pIC broadly tuned neuron responding to the same stimuli used in **a**. **c**, Duration and adaptation index for cool neurons in S1 (n = 122 cool transient neurons, 118 cool sustained neurons, 3 mice, 7 sessions).

Source data

Extended Data Fig. 5 Thermometer neurons.

a, Two-photon calcium imaging response amplitude plotted as a function of the thermal stimulus for neurons with similar response profile to neuron #5 of Fig. 4c (n = 16 thermometer neurons from 411 warm neurons in Fig. 4; ~5%). Grey lines show data for individual neurons, coloured filled circles show mean ± s.e.m. at AT 32 °C. Most thermometer neurons are characterized by low threshold response to warming stimuli and had cool responses only for small amplitude cooling stimuli. **b**, Grand average responses (mean ± s.e.m.) of thermometer neurons to 2 °C cooling (32-30 °C, blue) or 10 °C warming (32–42 °C, red) from AT 32 °C. Thermometer neurons conserve the latency difference as seen for cool and warm responses in tightly tuned and broadly tuned neurons (Fig. 3b, c). The latency of thermal responses in these neurons indicate that they receive the same input as the tightly tuned and broadly tuned neurons. These neurons highlight complex thermal features possibly constructed by the pIC. One hypothesis is that the lack of responses to large cooling stimuli could be explain by disynaptic, feedback, inhibition driven by a secondary cool channel activated by high amplitude cooling stimuli (eg. 32-24 °C).

Source data

Extended Data Fig. 6 Widefield imaging analysis of thermal amplitude encoding in pIC.

a, Left, schematic of widefield calcium imaging in a mouse implanted with a cranial window over pIC. **b**, Widefield responses of pIC to forepaw thermal stimulation with different amplitudes (n = 123–128 trials per temperature, 10 mice). Heat maps showing widefield responses to thermal stimuli with each line representing a single trial response and thermal stimuli below with white lines showing the onset and the end of the plateau phase of thermal stimuli; coloured lines to the right of heatmap show mean ± s.e.m. across trials from all mice and grey lines show mean response from individual mice; coloured shaded boxes indicate start of thermal stimulus and end of plateau phase. Right, same as left panels but for warm responses. **c**, Top, response latencies to different amplitudes of thermal stimuli from individual mice. Bottom, grey filled circles show data from all amplitudes (n = 10 mice). Filled circles show mean ± s.e.m. **d**, Top, peak population response to different amplitude stimuli from experiments presented in **b**. Coloured filled circles with connecting lines show data with AT of 32 °C. Grey filled circles with connecting dashed lines show data using AT of 26 °C (filled circles show mean ± s.e.m) (n = 6 mice). Green filled circles shows activity at AT. Bottom, same data as in top graph but showing population response amplitude plotted against thermal stimulus amplitude. Graph shows shift in warm response amplitude but not for cool when presented with a lower AT.

Source data

Extended Data Fig. 7 Electrophysiological validation of optogenetic inhibition and behavioural controls on S1 optogenetic silencing.

a, Example coronal slices from VGAT-ChR2 mice co-labelled with anti-GFP antibody showing homogeneous expression across pIC. Similar results were found in two other mice. **b**, Left, schematic showing an awake VGAT-ChR2 mouse with right forepaw on a Peltier element during electrophysiological recording and optogenetic manipulation of left pIC. Center, example coronal

section of a VGAT-ChR2 mouse showing 2 tracks of the 4 shanks electrode in magenta. Similar results were found in another mouse. Right, trial structure showing the timing of the optical stimulation stimulus relative to the thermal stimulus. **c**, Left, raster plot of example unit responses to 20 trials of cool (blue) or warm (red) stimulation and during thermal stimulation during optogenetic silencing of pIC (10 trials, cyan). Grey background box indicates timing of stimulus presentation for cool (blue) and warm (red). Pulsed LED on for the entire duration of the thermal stimulus. Thermal stimuli presentation was randomized. Trials with pIC optogenetic inhibition were also randomized, but are separated here for clarity. Right, shows graph of percentage reduction in firing rate of excitatory units during optogenetic stimulus trials in 4 mice, blue filed circles show 79 cold units and red shows 53 warm unit, black bars show mean ± s.e.m. **d**, Optical stimulation of the thermal zone of pIC in mice that do not express ChR2 (Thy1-GCaMP6s) does not affect cool perception. Behavioural task was identical to that presented in Fig. 5. Top graph shows number of trials with a lick during optical stimulation (cyan) vs. without optical stimulation (blue or red) at different thermal stimulus amplitudes, green shows catch trials. Grey thin lines show individual mice ($n = 5$), coloured filled circles connected by thick grey thick lines shows mean ± s.e.m. Lower panel shows the effect of light stimulus as the change in percentage of trials with licks, (mean ± s.e.m) corresponding to the stimulus amplitude above. **e**, same as in **d** but for warm. **f**, Top, proportion of trials with licks in VGAT-ChR2 mice ($n = 4$) during optical stimulation of S1 (cyan) vs. trials without optical stimulation (blue or red) for different cool stimulus amplitudes green shows catch trials. Grey lines show individual mice, coloured filled circles connected by thick grey lines show mean ± s.e.m. Middle, same as top but for simultaneous optogenetic inhibition of S1 and pIC. Bottom, shows the effect of light stimulus (change in percentage of trials with licks, mean ± s.e.m.) from mice with optogenetic inhibition of S1 alone (yellow filled circles) or from mice with optogenetic inhibition of both S1 and pIC simultaneously. **g**, Same as **e**, but for warm stimulation ($n = 6$ mice). S1 inactivation caused a smaller effect on cool perception in comparison to our previous report[7], however here we used a small diameter optical fiber for transient optogenetic inhibition, as compared to prior pharmacological inhibition.

Source data

Extended Data Fig. 8 Behavioural control experiments.

a, Optogenetic inhibition of pIC using the VGAT-ChR2 mice does not affect spontaneous licking behaviour. Left, trial structure showing optical stimulation during free licking (grey bar). Middle, example raster plot of free licking rates in a VGAT-ChR2 mouse during optical stimulation. Right, mean lick rates in VGAT-ChR2 mice during optical stimulation Off (black; 2 s before light stimulus) vs. On (cyan; 2 s beginning of light stimulus) trials. Grey thin lines show individual mice ($n = 6$), coloured filled circles connected by thick grey lines show mean ± s.e.m . **b**, Auditory perception is not altered by optogenetic inhibition of pIC. Perception task design is same as thermal task but uses an acoustic stimulus (14 kHz, 65 dB SPL). Left, structure of optical stimulation trial. Right, mean lick rates in VGAT-ChR2 mice during trials with optical stimulation during the auditory stimuli (cyan) vs. trials without optical stimulation (black). Grey thin lines show individual mice ($n = 4$), coloured filled circles connected by thick grey thick lines shows mean ± s.e.m. **c**, Left, schematic of discrimination task and placement of the optic fiber. Middle, example trial structure showing timing of reward window (grey) and the timing of an optical stimulus during trials with optogenetic manipulation. Filled circles show licks, first lick coloured to show rewarded lick. VGAT-ChR2 mice were trained to discriminate a 14 kHz acoustic stimulus (Go trial) from the same stimulus presented 0.6 s after the beginning of a 2 s thermal stimulus (cooling or warming) (NoGo trial). Right, response categories of the task. Go trials were rewarded, NoGo and catch trials (no stimuli) were not rewarded. Training continued until Go trial hit rate > 70%, NoGo false alarm rate ~50% and false alarm rate < 30%. A criterion of ~50% for the NoGo false alarm rate was chosen to be able to observe an increase or decrease in lick rates during optogenetic inhibition. The amplitude of the thermal stimuli (cooling or warming) was adjusted from mouse to mouse to obtain this criterion. **d**, Top, proportion of trials with licks in VGAT-ChR2 mice during optical stimulation (cyan) trials vs. without optical stimulation (black). Blue and red lines show data from individual mice trained with either cooling (blue, $n = 4$) or warming (red, $n = 3$) stimuli. Coloured filled circles joined by grey thick lines show mean ± s.e.m. Bottom, shows the effect of light stimulus (change in percentage of trials with licks, mean ± s.e.m.). The increase, rather than decrease, in NoGo

false alarm rate during optogenetic manipulation confirms the direct and selective involvement of pIC in thermosensory processing.

Source data

Supplementary information

Reporting Summary

Peer Review File

Source data

Source Data Figs. 1–5 and Source Data Extended Data Figs. 1–8.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Vestergaard, M., Carta, M., Güney, G. *et al.* The cellular coding of temperature in the mammalian cortex. *Nature* **614**, 725–731 (2023). https://doi.org/10.1038/s41586-023-05705-5

- Received: 14 January 2022
- Accepted: 04 January 2023
- Published: 08 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05705-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

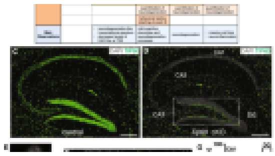
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

The brain’s encoding of warm and cool

Nature Research Briefing 08 Feb 2023

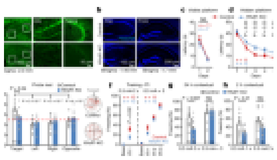
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05705-5



TIP60/KAT5 is required for neuronal viability in hippocampal CA1

07 November 2019

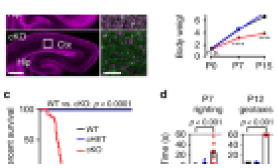
Inga Urban, Cemil Kerimoglu, … Gregor Eichele



m6A facilitates hippocampus-dependent learning and memory through YTHDF1

31 October 2018

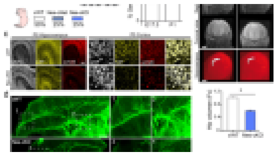
Hailing Shi, Xuliang Zhang, … Tao Zhou



Deletion of Topoisomerase 1 in excitatory neurons causes genomic instability and early onset neurodegeneration

23 April 2020

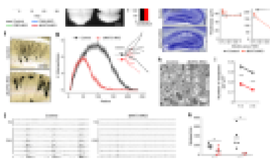
Giulia Fragola, Angela M. Mabb, … Mark J. Zylka



[CBP and SRF co-regulate dendritic growth and synaptic maturation]

08 March 2019

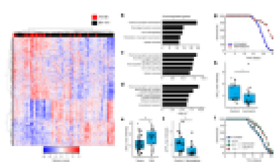
Beatriz del Blanco, Deisy Guiretti, … Angel Barco



[KAT3-dependent acetylation of cell type-specific genes maintains neuronal identity in the adult mouse brain]

22 May 2020

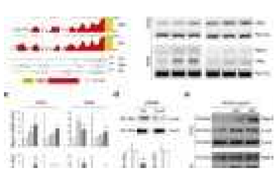
Michal Lipinski, Rafael Muñoz-Viana, … Angel Barco



[Regulation of lifespan by neural excitation and REST]

16 October 2019

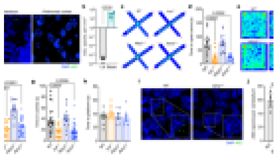
Joseph M. Zullo, Derek Drake, … Bruce A. Yankner



[Wnt/β-catenin signaling stimulates the expression and synaptic clustering of the autism-associated Neuroligin 3 gene]

05 March 2018

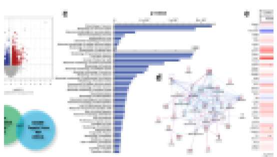
Matías A. Medina, Víctor M. Andrade, … Giancarlo V. De Ferrari



[AIM2 inflammasome surveillance of DNA damage shapes neurodevelopment]

08 April 2020

Catherine R. Lammert, Elizabeth L. Frost, … John R. Lukens



[KHSRP loss increases neuronal growth and synaptic transmission and alters memory consolidation through RNA stabilization]

07 July 2022

Sarah L. Olguin, Priyanka Patel, … Nora I. Perrone-Bizzozero

- Article
- Open Access
- Published: 15 February 2023

A NPAS4–NuA4 complex couples synaptic activity to DNA repair

- Elizabeth A. Pollina,
- Daniel T. Gilliam,
- Andrew T. Landau,
- Cindy Lin,
- Naomi Pajarillo,
- Christopher P. Davis,
- David A. Harmin,
- Ee-Lynn Yap,
- Ian R. Vogel,
- Eric C. Griffith,
- M. Aurel Nagy,
- Emi Ling,
- Erin E. Duffy,
- Bernardo L. Sabatini,
- Charles J. Weitz &
- …
- Michael E. Greenberg

Show authors

Nature **volume 614**, pages 732–741 (2023)

- 14k Accesses

- 191 Altmetric

- Metrics details

Abstract

Neuronal activity is crucial for adaptive circuit remodelling but poses an inherent risk to the stability of the genome across the long lifespan of

postmitotic neurons[1,2,3,4,5]. Whether neurons have acquired specialized genome protection mechanisms that enable them to withstand decades of potentially damaging stimuli during periods of heightened activity is unknown. Here we identify an activity-dependent DNA repair mechanism in which a new form of the NuA4–TIP60 chromatin modifier assembles in activated neurons around the inducible, neuronal-specific transcription factor NPAS4. We purify this complex from the brain and demonstrate its functions in eliciting activity-dependent changes to neuronal transcriptomes and circuitry. By characterizing the landscape of activity-induced DNA double-strand breaks in the brain, we show that NPAS4–NuA4 binds to recurrently damaged regulatory elements and recruits additional DNA repair machinery to stimulate their repair. Gene regulatory elements bound by NPAS4–NuA4 are partially protected against age-dependent accumulation of somatic mutations. Impaired NPAS4–NuA4 signalling leads to a cascade of cellular defects, including dysregulated activity-dependent transcriptional responses, loss of control over neuronal inhibition and genome instability, which all culminate to reduce organismal lifespan. In addition, mutations in several components of the NuA4 complex are reported to lead to neurodevelopmental and autism spectrum disorders. Together, these findings identify a neuronal-specific complex that couples neuronal activity directly to genome preservation, the disruption of which may contribute to developmental disorders, neurodegeneration and ageing.

Main

Sensory experience is essential for proper neuronal maturation and circuit plasticity[1]. The signalling cascades initiated by experience-driven neuronal activity culminate in the induction of gene programmes that control diverse processes such as dendrite and synapse growth, synapse elimination, recruitment of inhibitory neurotransmission, and adaptive myelination[6,7]. However, neuronal activity also threatens the genomic integrity of postmitotic neurons that must survive the lifetime of an organism. For example, heightened metabolic demands during periods of elevated activity may increase oxidative damage to actively transcribed regions of the genome[8]. Activity-induced transcription itself poses a further threat to genome stability, as it has been linked to the induction of repeated DNA double-strand breaks (DSBs) at regulatory elements, such as the promoters

of stimulus-inducible genes[2,3,4,5,9,10]. Although the coupling of transcription to DNA breaks is observed across cell types, this process poses a specific challenge to long-lived neurons, which cannot use replication-dependent DNA repair pathways and possess limited regenerative mechanisms to replace damaged cells[11]. Accumulating DNA damage to neuronal genomes is a cardinal feature of neurodegenerative disorders and organismal ageing[12,13]. Thus, understanding the strategies that neurons use to prevent and repair damage may have direct translation to human longevity and ageing therapies. So far, there are no examples of neuronal-specific repair machinery that mitigate the risks of genome instability during heightened activity. By investigating features of the activity-dependent transcriptional programme specific to neurons, we discover a biochemical coupling of neuronal activity to DNA repair through a previously unknown form of the NuA4 chromatin remodeller–DNA repair complex that assembles around the inducible, neuronal-specific transcription factor NPAS4.

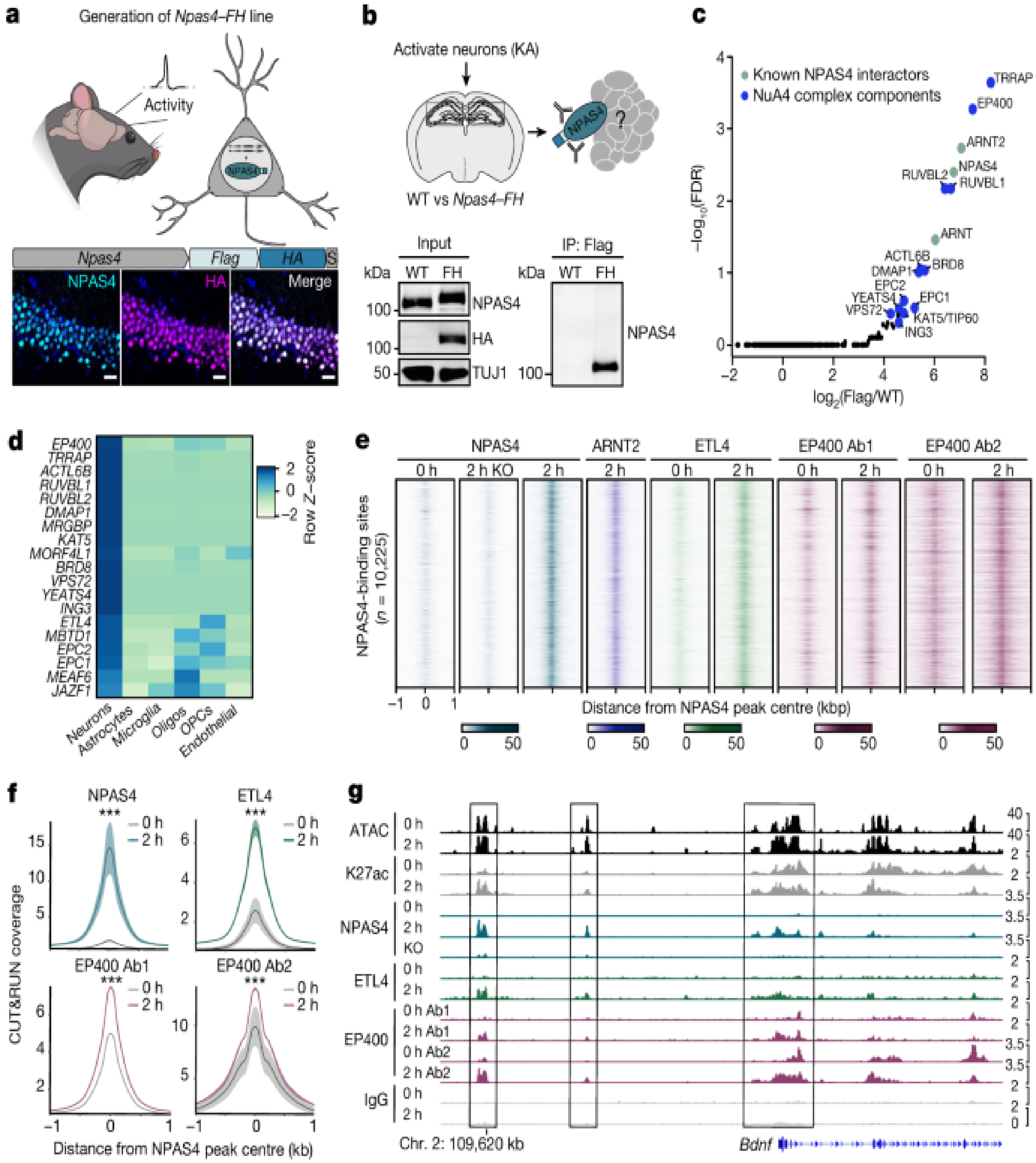
Identification of the NPAS4–NuA4 complex

Unlike most activity-inducible transcription factors, which are broadly expressed and induced by various stimuli, NPAS4 is selectively expressed in neurons following membrane depolarization-induced calcium signalling[14]. To understand the functions of this factor, which is specifically attuned to neuronal activity, we sought to purify NPAS4-containing protein complexes from the adult mouse brain. We reasoned that NPAS4 might assemble into a multisubunit complex that expands its biochemical activities in activated neurons. Using size-exclusion chromatography and non-denaturing gel electrophoresis, we observed that NPAS4 resides in a high molecular weight complex of around 1 MDa. As the predicted size of NPAS4 with either of its heterodimer partners (ARNT1 and ARNT2) is around 175 kDa, this finding suggests that NPAS4 interacts with multiple unknown protein partners (Extended Data Fig. 1a).

To facilitate the purification of this putative NPAS4 complex, we generated *Npas4–Flag-HA* and *Arnt2–Flag-HA* knock-in mouse lines in which the epitope tags Flag and haemagglutinin (HA) are appended to the carboxy termini of NPAS4 and ARNT2. In homozygous knock-in mice, we validated the correct genomic insertion of the tags and verified overlapping

immunostaining of endogenous NPAS4 or ARNT2 and the HA epitope (Fig. 1a,b and Extended Data Fig. 1b,c). We also demonstrated that wild-type and tagged NPAS4–Flag-HA (NPAS4–FH) exhibit similar expression levels and induction kinetics (Extended Data Fig. 1d). To generate high levels of NPAS4 required for biochemical purification, we stimulated neurons in the hippocampus of *Npas4–FH* mice through the injection of low-dose kainic acid (KA), a glutamate receptor agonist that synchronously depolarizes hippocampal neurons. We then immunopurified NPAS4 using anti-Flag antibodies and performed mass spectrometry (Fig. 1c and Extended Data Fig. 1e). The mass spectrometry data revealed interactions between NPAS4 and all reported subunits of a single chromatin modifier, the NuA4 complex, the estimated size of which is approximately 1.0–1.3 MDa (Fig. 1c and Supplementary Table 1)[15,16,17].

Fig. 1: Neuronal activity assembles the NPAS4–NuA4 complex on chromatin.



a, Top, schematic of generation of the *Npas4–FH* mouse model. Bottom, representative immunohistochemistry (performed in triplicate) for NPAS4 and HA in hippocampus samples from *Npas4–FH* mice following 2 h of KA stimulation. Scale bars, 25 µm. **b**, NPAS4–FH and associated protein complexes isolated by anti-Flag immunoprecipitation (IP). Representative western blot (performed in triplicate) with anti-NPAS4 antibody confirms immunoprecipitation only in *Npas4–FH* mice. For gel source data, see Supplementary Fig. 1. TUJ1 processing control was run on a separate gel. **c**,

NPAS4-interacting proteins. $-\log_{10}$(false discovery rate (FDR)) compared with $\log_2$-transformed fold change in peptides obtained by anti-Flag immunoprecipitation in stimulated *Npas4–FH* tissue compared with wild-type tissue ($n = 3$ pools of 8–10 mice), followed by mass spectrometry. Green points indicate known NPAS4 interactors. Blue points indicate NuA4 components. **d**, Expression levels of NuA4 subunits across cell types in human primary motor cortex, displayed as row *Z*-score. snRNA-seq data published by the Allen Brain Institute[18]. Oligos, oligodendrocytes; OPCs, oligodendrocyte precursor cells. **e**, CUT&RUN signals for NPAS4–NuA4 components in either unstimulated (0 h) or stimulated (2 h) hippocampal tissue. Each NPAS4-binding site is represented as a horizontal line centred at the peak summit and extended ±1 kb. The colour intensity represents the read depth (normalized to 10 million) as indicated by the scale bar for each factor. Data plotted show the aggregate signal from all replicates per factor. Ab1, antibody 1 (Bethyl A300-541-A); Ab2, antibody 2 (Abcam Ab5201). For **e**–**g**, $n = 3$–5 mice pooled per replicate. Replicate numbers provided in Supplementary Table 2. **f**, Aggregate CUT&RUN coverage (fragment depth per bp per peak) at NPAS4-binding sites. Signals displayed as mean ± s.e.m. ***$P < 2.2 \times 10^{-16}$. *P* values were calculated from the average signal extracted in a 2 kb window centred on NPAS4 peaks using unpaired, two-tailed Wilcoxon rank-sum tests. **g**, Integrative Genomics Viewer (IGV) tracks of NPAS4–NuA4 CUT&RUN signal at the activity-inducible gene *Bdnf*. Data plotted show the aggregate signal from all replicates per factor. The *y* axis displays normalized coverage scaled for each factor at the chosen locus.

Source Data

We first demonstrated co-immunoprecipitation between NPAS4 and several NuA4 subunits (TRRAP, EP400 and DMAP1) in the visual cortex using light exposure as a physiological stimulus to induce neuronal activity (Extended Data Fig. 1f). We further characterized this complex by immunoprecipitating either NPAS4 or a component of the NuA4 complex, TIP60 (also known as KAT5), followed by immunoblotting and mass spectrometry analyses. These experiments confirmed that the interaction between NPAS4 and the NuA4 complex is reciprocal. Moreover, NPAS4–ARNT2 dimers were among the most abundant transcription factors

associated with NuA4 in the brain. In addition, a new subunit of the complex, the poorly characterized protein ETL4, was identified (Extended Data Figs. 1g–j and 2a,b and Supplementary Table 1). We observed that several subunits of the NuA4 complex interacted before stimulation, which suggests that following NPAS4 induction, NPAS4–ARNT dimers join a pre-existing complex (Extended Data Fig. 2b). However, we cannot exclude the possibility that other subunits associate with, or post-translational modifications are added to, the NuA4 complex following activity. NPAS4 is the major inducible component of the complex at the RNA level (Extended Data Fig. 2c). Neither the related protein NPAS3 nor another activity-inducible factor, FOS, co-immunoprecipitated with NuA4 components in the brain. This result highlights the specificity of the NPAS4–NuA4 interaction (Extended Data Fig. 2b). Moreover, NPAS4, but not other activity-inducible transcription factors such as FOS and EGR1, interacted with NuA4 components in a heterologous expression system (HEK293T cells) (Extended Data Fig. 2d,e). In both human and mouse cells, expression of these 19 new interactors of NPAS4 were enriched in neurons compared with other brain cell types[18] (Fig. 1d and Extended Data Fig. 2f,g). This result suggests that within the brain, the NPAS4–NuA4 complex functions specifically within neurons.

To determine whether NPAS4 and NuA4 co-localize on chromatin, we performed CUT&RUN[19] to obtain a map of NPAS4–NuA4 genomic binding in stimulated hippocampal nuclei isolated from mice injected with low doses of KA (Fig. 1e and Supplementary Table 2). We confirmed the specificity of the NPAS4 CUT&RUN signal by performing NPAS4 CUT&RUN on nuclei isolated from unstimulated brains (in which little to no NPAS4 is expressed) and on nuclei isolated from brains of stimulated *Npas4* knockout mice. This experiment generated a list of 10,225 high-confidence NPAS4-binding sites (Fig. 1e and Extended Data Fig. 3a–c). These binding sites were highly correlated with NPAS4 signal from chromatin immunoprecipitation assays with sequencing (ChIP–seq) and showed enrichment of the E-box and bHLH–PAS binding motifs (Extended Data Fig. 3d,e). NPAS4, its partner ARNT2, the NuA4 component EP400 and the newly identified NuA4 subunit ETL4 also co-localized across the genome in stimulated neurons (Fig. 1e–g and Extended Data Fig. 3f,g). Moreover, binding of both NPAS4 and ETL4 to the genome was highly inducible at NPAS4 sites (Fig. 1f). By

contrast, EP400 was present at NPAS4-binding sites before stimulation, which suggests that it may be retained at these sites in the absence of NPAS4 owing to the ability of NuA4 to bind to acetylated histones[20] or by NuA4-independent binding of EP400 (Fig. 1f,g). However, significantly more EP400 CUT&RUN signal was observed at NPAS4 sites that lack FOS than the converse (that is, sites with FOS but no NPAS4). This result demonstrates the specificity of NPAS4 and EP400 co-binding rather than a general recruitment of EP400 by activity-inducible transcription factors (Extended Data Fig. 3h,i). In summary, our biochemical evidence and genomic binding assays identified a neuronal-specific form of the NuA4 complex that assembles with NPAS4 in multiple brain regions in an activity-dependent manner.

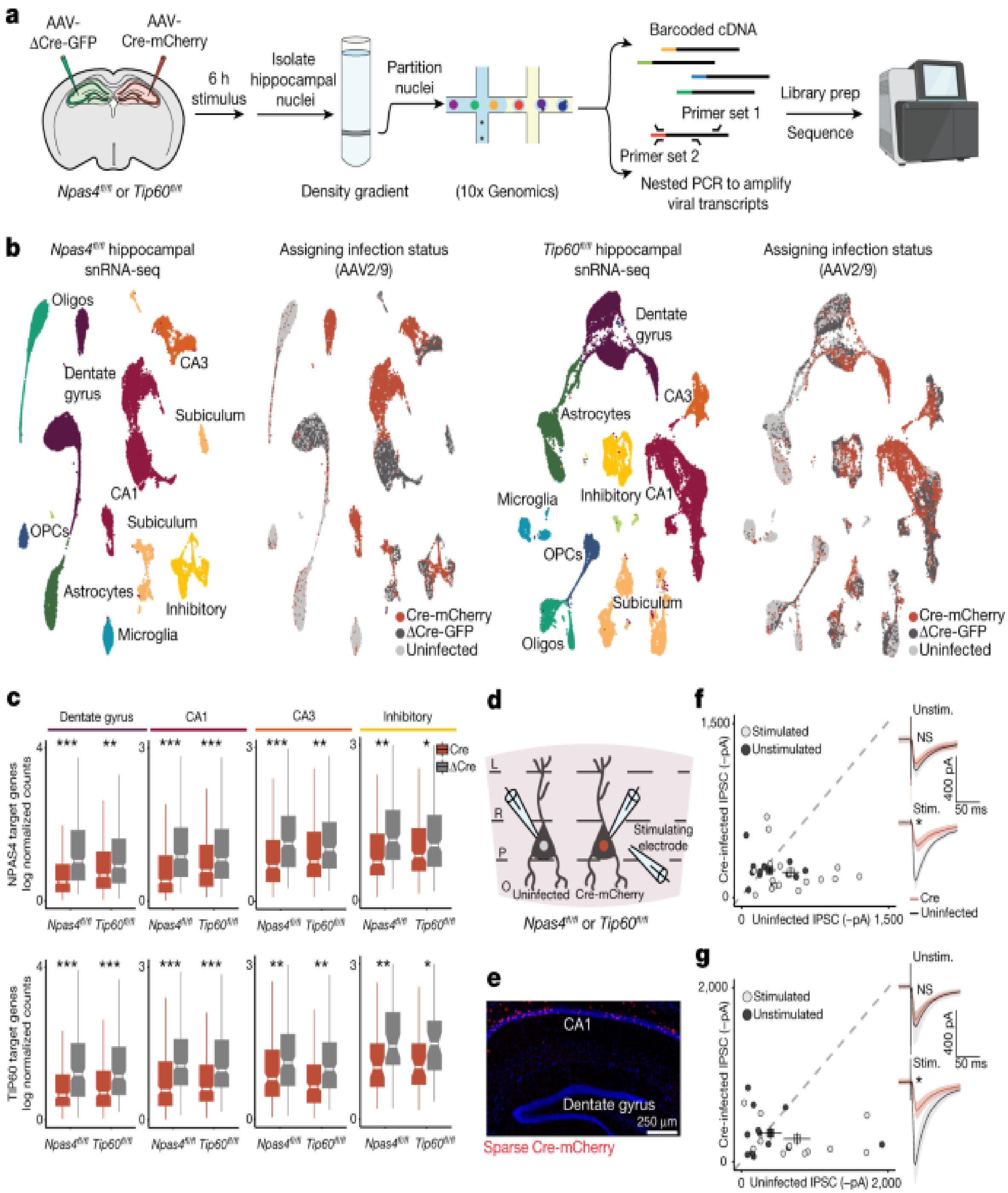
NPAS4–NuA4-driven inducible gene programmes

We next investigated the functions of the NPAS4–NuA4 complex in the brain. Studies using yeast, flies and non-neuronal cells have ascribed two activities to the NuA4 complex: (1) controlling transcription through chromatin regulation[15,16,17] and (2) coordinating repair of DNA DSBs[21,22,23,24,25,26]. We therefore considered the possibility that NPAS4, by recruiting NuA4, might serve a dual purpose. Previous studies of NPAS4 suggest that the NPAS4–NuA4 complex probably has a central role in activity-dependent gene regulation[14,27]. In addition, the role of NuA4 in DSB repair suggests that NPAS4 might have a previously undescribed function in neuronal activity-dependent DNA repair at promoters and enhancers. We examined these two possible functions of NPAS4–NuA4 by disrupting either NPAS4 or TIP60, the histone acetyltransferase component of NuA4 (refs. [15,16]).

To determine whether NPAS4 and NuA4 coordinate activity-dependent transcription, we injected *Npas4*$^{fl/fl}$ or *Tip60*$^{fl/fl}$ mice with a virus expressing either Cre-mCherry or a recombination-deficient version of Cre (ΔCre-GFP) into contralateral sides of the hippocampus to unilaterally remove either *Npas4* or *Tip60* (Fig. 2a and Extended Data Fig. 4a–c). Note that depletion of *Tip60* does not impair the induction of NPAS4 (Extended Data Fig. 4d,e). We administered a low dose of KA to activate all classes of neurons present in the hippocampus, collected nuclei 6 h later to capture

maximal induction of NPAS4 target genes[14] and performed single-nucleus RNA-sequencing (snRNA-seq) (Fig. 2a and Extended Data Fig. 4f–i). In both *Tip60*$^{fl/fl}$ and *Npas4*$^{fl/fl}$ datasets, we identified ten principal cell types, including five neuronal subtypes (dentate gyrus, CA1, CA3, subiculum and inhibitory), and multiple non-neuronal cell types. We used a nested PCR strategy as an additional step in library preparation to amplify viral transcripts. This step enabled us to assign infection status to each nucleus as Cre-infected (cKO), ΔCre-infected (wild-type) or uninfected (Fig. 2a,b). In both datasets, within each neuronal cluster, the cells subclustered according to infection status (Cre compared with ΔCre). This effect was not driven by the expression of viral transcripts (Methods), but rather reflects the inability of Cre-infected cells to fully induce activity-dependent genes following the depletion of either NPAS4 or TIP60.

Fig. 2: NPAS4–NuA4 regulates activity-dependent transcription and recruitment of somatic inhibition.



a, Scheme of the experimental design. Nuclei were isolated from hippocampi of *Npas4*$^{fl/fl}$ and *Tip60*$^{fl/fl}$ mice expressing either Cre-mCherry or ΔCre-GFP, followed by snRNA-seq. An additional PCR step to amplify viral transcripts within the cDNA library was used to assign infection status to each nucleus. **b**, Left, uniform manifold approximation and projection (UMAP) visualizations of *Npas4*$^{fl/fl}$ and *Tip60*$^{fl/fl}$ snRNA-seq datasets. Right,

UMAP visualizations of infection status. *Npas4*$^{fl/fl}$: 32,418 nuclei from 2 mice, 12,963 Cre-infected, 8,845 ΔCre-infected. *Tip60*$^{fl/fl}$: 44,511 nuclei from 3 mice, 13,536 Cre-infected, 13,461 ΔCre-infected. **c**, Boxplots showing the average expression (Seurat log(e) normalized counts) of NPAS4 or TIP60 target genes (Methods) comparing Cre-infected and ΔCre-infected nuclei within each neuronal celltype. Boxplots show the median (line), inter-quartile range (IQR; box) and 1.5× IQR (whiskers), and notches indicate the median ± 1.58× IQR/sqrt(n). ***$P < 5 \times 10^{-10}$, **$P < 5 \times 10^{-4}$, *$P < 0.01$. P values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests. Exact P values and cell numbers per cluster provided in the source data. **d**, Schematic of the recording configuration, showing the stimulating electrode in the centre of the stratum pyramidale to measure IPSCs onto neighbouring Cre-infected and uninfected CA1 pyramidal neurons in either *Npas4*$^{fl/fl}$ or *Tip60*$^{fl/fl}$ acute hippocampal slices. L, lacunosum; O, oriens; P, pyramidale; R, radiatum. **e**, Representative image (performed in triplicate) showing sparse Cre-mCherry expression in the CA1. **f**,**g**, Scatterplots of IPSC amplitudes recorded from pairs of neighbouring uninfected and Cre-infected cells in unstimulated (Unstim.; saline) or stimulated (Stim.; low-dose KA) *Npas4*$^{fl/fl}$ (**f**) or *Tip60*$^{fl/fl}$(**g**) mice. Points represent an uninfected–Cre-infected pair. Squares represent the average ± s.e.m. Right insets, traces of uninfected or Cre-infected cells. *Npas4*$^{fl/fl}$: stimulated: $n = 16$ pairs from 4 mice; unstimulated: $n = 12$ pairs from 5 mice. *Tip60*$^{fl/fl}$: stimulated: $n = 11$ pairs from 2 mice; unstimulated: $n = 12$ pairs from 2 mice. *$P = 3.73 \times 10^{-3}$ for *Npas4*$^{fl/fl}$; *$P = 0.0209$ for *Tip60*$^{fl/fl}$. Data are mean ± s.e.m. P values were calculated using unpaired, two-tailed t-tests. The sequencer image in **a** is from BioRender (https://biorender.com).

Source Data

We identified differentially expressed genes (Methods; negative fold change (Cre/ΔCre); adjusted $P < 0.01$) within each principal neuronal subtype and focused our analysis on downregulated genes that are more likely to be direct targets of the complex. Across each of the neuronal subtypes, genes downregulated owing to *Npas4* deletion were also significantly downregulated following *Tip60* loss in the corresponding cell cluster of the *Tip60*$^{fl/fl}$ dataset, with the reciprocal comparison producing the same result (Fig. 2c and Extended Data Fig. 5). In addition to capturing known NPAS4

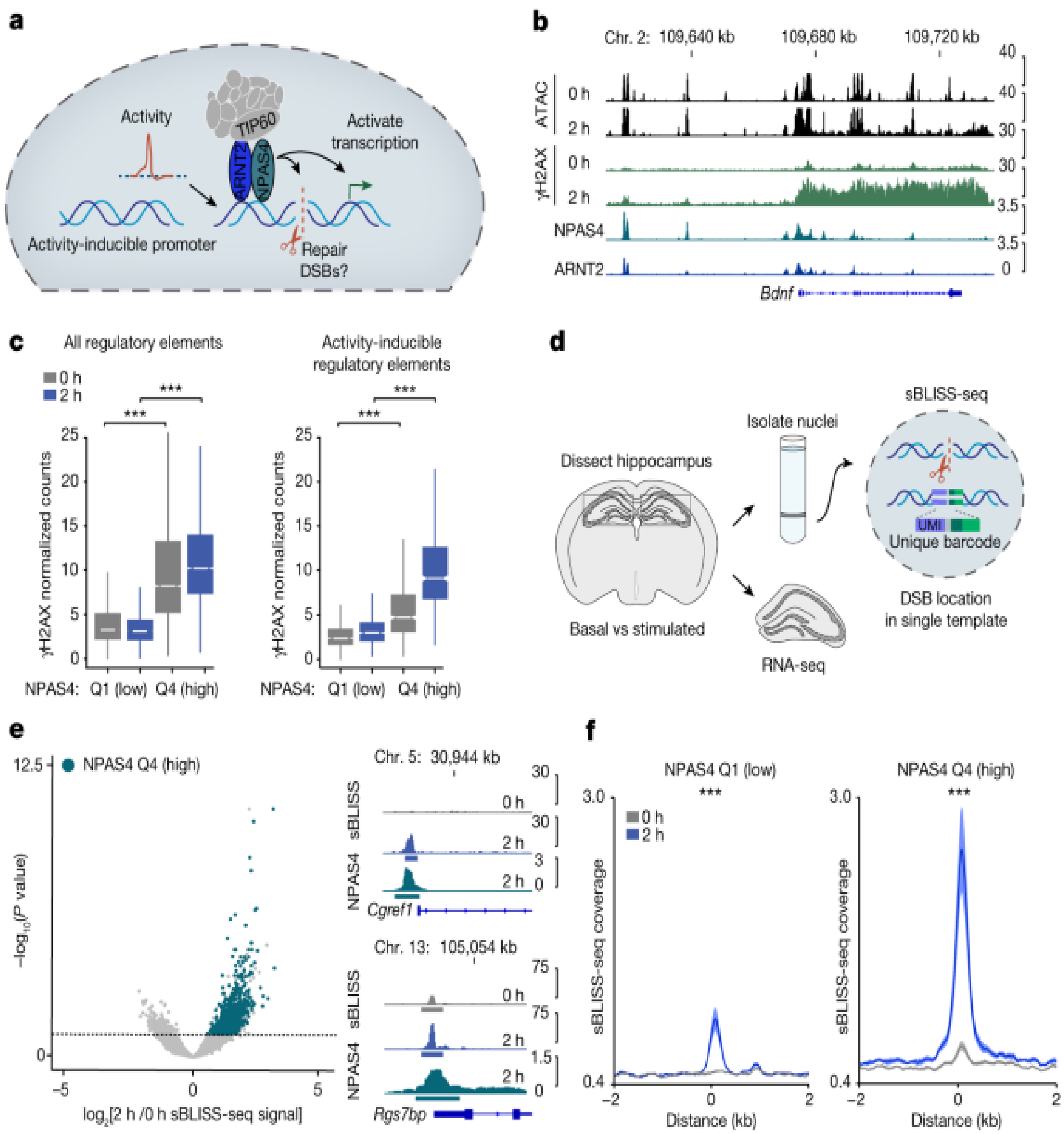
target genes such as *Nptx2*, *Plk2* and *Bdnf*, we identified 1,766 new targets of NPAS4 in the hippocampus and defined their cell-type-specific expression patterns (Supplementary Table 3). In primary cultures of mouse neurons, we independently confirmed that components of the NuA4 complex, NPAS4, TIP60 and EP400, each regulate the same inducible transcriptional programme (Extended Data Fig. 6a–c). As further corroboration of the importance of the NPAS4–NuA4 interaction for gene activation, expression of truncated forms of NPAS4 that do not strongly interact with NuA4 (Extended Data Fig. 2d) significantly impaired the ability of NPAS4 to activate its target binding sites in luciferase reporter assays (Extended Data Fig. 6d,e). Together, these findings provide evidence that NPAS4 and NuA4 are not only key regulators of neuronal activity-dependent gene transcription but also activate gene expression as a single functional unit.

As a final test confirming that NPAS4 and NuA4 function together as a complex in neurons, we asked whether TIP60 has the same role in activated neuronal circuits as observed for NPAS4. A key function of NPAS4 in excitatory CA1 pyramidal neurons is to mediate recruitment of somatic inhibition in hippocampal pyramidal neurons in response to neuronal activity[7,28]. We therefore sparsely injected the CA1 region of *Npas4*$^{fl/fl}$ and *Tip60*$^{fl/fl}$ mice with an adeno-associated virus (AAV) encoding Cre-mCherry and performed simultaneous recordings of evoked inhibitory postsynaptic currents (IPSCs) from neighbouring wild-type (Cre-mCherry^{-}) and cKO (Cre-mCherry^{+}) pyramidal neurons in acute hippocampal slices (Fig. 2d,e). Axons within the stratum pyramidale were stimulated with the minimum strength required to generate IPSCs from wild-type and cKO neurons. Deletion of NPAS4 or TIP60 significantly reduced activity-dependent somatic inhibition induced by low-dose KA stimulation. By contrast, no differences between Cre-infected and uninfected cells were observed in saline-injected control animals (Fig. 2f,g and Extended Data Fig. 6f–i). Together, these experiments indicate that the NPAS4–NuA4 complex assembles in activated neurons to coordinate inducible gene transcription and to dynamically reorganize stimulated neuronal circuits in the brain.

Inducible DNA breaks at NPAS4–NuA4 sites

The conserved function of NuA4 in DSB repair[21,22,23,24,25,26] prompted us to next investigate the possibility that NPAS4, through integration into the NuA4 complex, might play a previously unrecognized role in DNA damage control in neurons. Notably, stimulus-dependent gene induction in neurons has been suggested to result in DSBs, which are probably mediated by topoisomerase enzymes such as TOP2B, as paused RNA Pol II is released into productive elongation[3,4,9,10,29]. However, the extent to which DSBs occur in response to neuronal activity in vivo and the mechanisms that neurons might use to repair these breaks and mitigate accumulating damage remain unclear. We proposed that the formation of a complex between NPAS4 and NuA4 may represent a mechanism by which neurons efficiently drive activity-induced transcriptional responses while simultaneously preserving genome stability downstream of neuronal activity (Fig. 3a). We sought to identify genomic loci that undergo both damage and repair in response to activity. We also examined whether these regions are targeted by NPAS4–NuA4 and tested whether perturbing NPAS4–NuA4 impairs DNA repair at these sites.

Fig. 3: NPAS4–NuA4-bound sites undergo recurrent DNA breaks in vivo.



a, A model for the dual function of NPAS4–NuA4 in stimulating transcription and DNA repair in active neurons. **b**, IGV tracks displaying aggregate signals for ATAC-seq ($n = 3$), NPAS4 CUT&RUN ($n = 5$), ARNT2 CUT&RUN ($n = 2$) and γH2AX ChIP–seq ($n = 3$) at the activity-inducible gene *Bdnf*. $n = 3$–5 mice pooled per replicate. **c**, Boxplots of average γH2AX ChIP–seq normalized counts ($n = 3$ pools of 3–5 mice) at all regulatory elements (left) and activity-inducible regulatory elements (right), subset by quartiles of NPAS4 CUT&RUN signal (Methods). All regulatory elements: Q1 = 44,864 sites, Q4 = 7,378 sites. Activity-inducible elements:

Q1 = 1,017 sites, Q4 = 764 sites. Boxplots show the median (line), IQR (box) and 1.5× IQR (whiskers), and notches indicate the median ± 1.58× IQR/sqrt(*n*). ***$P < 2.2 \times 10^{-16}$. *P* values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests. **d**, Schematic of sBLISS-seq to map DSBs in brain nuclei. A tissue aliquot from each sBLISS-seq sample was kept for paired RNA-seq analysis. **e**, Volcano plot depicting the DeSeq2 $\log_2$-transformed fold change versus $-\log_{10}$(Benjamini–Hochberg adjusted *P* value) of sBLISS-seq signals between 0 and 2 h KA stimulation. Dotted line indicates adjusted $P < 0.1$. Elements with $\log_2$-transformed fold change > 0, adjusted $P < 0.1$ and that are within Q4 of NPAS4 IgG-normalized CUT&RUN signals are indicated in blue. Right inset, IGV tracks of aggregate sBLISS-seq ($n = 8$ individual mice) and NPAS4 CUT&RUN ($n = 5$ pools of 3–5 mice) at *Cgref1* and *Rgs7bp* promoters. Coloured bars represent statistically defined peaks. **f**, Aggregate plots showing sBLISS-seq coverage (fragment depth per bp per peak) at activity-inducible regulatory elements, subset by quartiles of NPAS4 binding. Signals are mean ± s.e.m. ($n = 8$ individual mice). ***$P < 2.2 \times 10^{-16}$. *P* values were calculated using the average signal extracted in a 500 bp window around the element centre using unpaired, two-tailed Wilcoxon rank-sum tests.

We used assay for transposase-accessible chromatin with sequencing (ATAC-seq) and CUT&RUN for the histone modification H3K27ac to identify the landscape of constitutive and activity-responsive genomic regulatory elements in the hippocampus. We defined regulatory elements as regions of the genome with either an ATAC-seq or H3K27ac CUT&RUN peak, hereafter referred to as ‘all regulatory elements’. We further defined ‘activity-inducible regulatory elements’ as sites that exhibited dynamic increases in the ATAC-seq signal (twofold increase; adjusted $P < 0.05$) and/or an increase in the H3K27ac CUT&RUN signal (1.5-fold increase; adjusted $P < 0.05$) after 2 hours of low dose KA stimulation (Extended Data Fig. 7a–c and Supplementary Table 4). Within this dataset, we classified NPAS4-bound elements as those that overlapped with our list of reproducible NPAS4 peaks. We first asked whether neuronal stimulation in vivo leads to chromatin signatures of DNA damage, especially at elements bound by NPAS4–NuA4. We performed ChIP–seq for the DNA damage-associated histone modification γH2AX (phosphorylated Ser139 on histone H2AX) in both unstimulated and stimulated neurons (Fig. 3b and Extended

Data Fig. 7d,e). After stimulation, we observed increased γH2AX levels at NPAS4-bound sites, with the maximal signal and inducibility at sites in the highest quartile of NPAS4 binding (Fig. 3c). Notably, we observed a larger increase in γH2AX levels following stimulation at NPAS4-bound elements that fell within our landscape of activity-inducible elements (Fig. 3c, right). Consequently, we focused primarily on these activity-inducible regulatory elements as they appeared to best capture the relationship between activity-induced transcription, NPAS4–NuA4 binding and DNA damage.

Next we directly mapped DSBs in vivo using suspension breaks in situ ligation and sequencing (sBLISS-seq), a sequencing-based method that identifies DSBs through the ligation of DNA sequencing adapters onto free DNA ends[30]. After validating the ability of sBLISS-seq to capture CRISPR–Cas9-induced DSBs (Extended Data Fig. 7f–h), we profiled the landscape of DSBs that occur in the adult hippocampus under basal conditions or following stimulation. In parallel, we performed RNA-seq on the same samples to examine how DSBs correlate with transcriptional dynamics (Fig. 3d). We first examined basic features of these DNA breaks across the genome at all time points. As previously reported[30], we observed maximal sBLISS-seq signals at the promoters of the most highly expressed genes (Extended Data Fig. 7i–k). There was also a significant correlation between the sBLISS-seq signal and the γH2AX ChIP–seq signal at 2 h after stimulation. This result provides an independent demonstration of damage to these sites (Extended Data Fig. 7l). Motifs enriched in statistically defined peaks of reproducible sBLISS-seq signal included several activity-dependent transcription factors such as ATF1, EGR and NPAS4–ARNT, which indicates that a subset of these breaks is driven by transcriptional induction downstream of increased neuronal activity (Extended Data Fig. 7m).

We next examined how neuronal stimulation influences the landscape of DSBs. Although neuronal activity did not alter the overall distribution of sBLISS-seq signal (DSBs) across the genome (Extended Data Fig. 7i), we identified 1,581 regulatory elements (adjusted $P < 0.1$) that significantly increased in DSB signal at 2 h after stimulation (Fig. 3e). The activity-inducible elements displayed increased DSB signals after stimulation, which was in contrast to the non-inducible elements (Extended Data Fig. 7n). Moreover, 69% of the elements that had significantly increased DSB signal

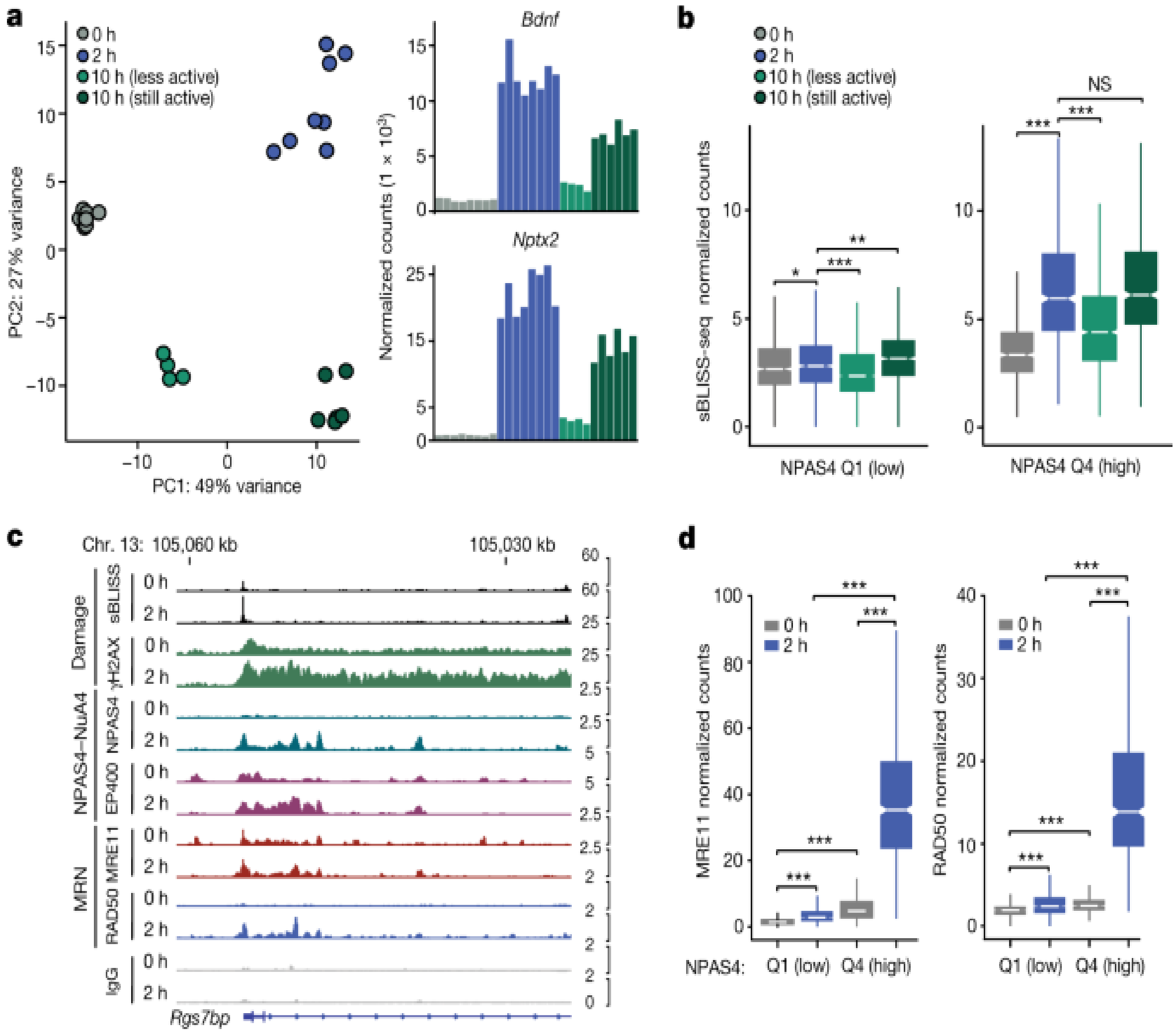
at 2 h after stimulation were in the top quartile of NPAS4 CUT&RUN signal, and the NPAS4-bound activity-inducible regulatory elements displayed the most significant increases in DSB signal after stimulation (Fig. 3e,f). We corroborated these findings using a complementary assay (END-seq)[31] to map DSBs in primary mouse cortical neurons at either 0 h or 2 h after stimulation with 55 mM KCl. To enhance sensitivity, we performed these assays in the presence of etoposide, which blocks the re-ligation of DSBs generated by topoisomerase enzymes. As reported previously[31], we observed enrichment of END-seq signal at CTCF-bound sites (Extended Data Fig. 8a). In line with our sBLISS-seq data from the hippocampus, the END-seq results revealed a stimulus-dependent increase in DNA breaks at NPAS4-bound sites in primary neurons (Extended Data Fig. 8b). Notably, the overall level and inducibility of both sBLISS-seq and END-seq signals were higher at sites that have a CUT&RUN peak for NPAS4 but not for the activity-dependent factor FOS than the converse (Extended Data Fig. 8c–i). Together, our independent measures of DNA damage by γH2AX ChIP–seq, sBLISS-seq and END-seq demonstrate that NPAS4 preferentially binds to sites that undergo activity-inducible DNA breaks in neurons.

NPAS4–NuA4 sites undergo repair

We next asked whether these damaged elements, particularly those bound by NPAS4–NuA4, also undergo repair. To this end, we examined the levels of DSBs at a third time point, 10 h after KA stimulation, reasoning that a subset of sites may return towards the baseline DSB signal as activity-induced transcription subsides. We therefore used RNA-seq datasets collected from the same tissue as the sBLISS-seq datasets to identify samples in which activity-dependent transcription was returning to baseline. Principal component analysis (PCA) and hierarchical clustering of the RNA-seq data revealed two clusters of samples at the 10 h time point. One cluster displayed a significantly lower level of activity-inducible gene expression than at the 2 h time point (termed 'less active'), whereas the other 10 h cluster maintained higher levels of inducible genes (termed 'still active') (Fig. 4a and Extended Data Fig. 9a,b). PCA of the sBLISS-seq data also demonstrated that these less-active samples clustered closer to unstimulated samples (Extended Data Fig. 9c). Notably, we observed a separation between unstimulated and the less-active 10-h samples, which suggests that

the less-active samples were returning to the baseline state rather than failing to initially stimulate. In the less-active 10-h samples, DSB signal at activity-inducible regulatory elements was significantly reduced relative to the 2 h time point. This result suggests that there was ongoing repair that resolves DSBs alongside declining transcriptional activity. This increase in the DNA break signal at 2 h after stimulation, coupled with a decrease at 10 h, was most pronounced at activity-inducible regulatory elements with the highest levels of NPAS4 binding (Fig. 4b and Extended Data Fig. 9d,e). We confirmed that NPAS4-bound sites undergo active DNA repair by examining published maps of DNA synthesis-dependent repair in human neurons (SAR-seq)[32]. Although this method does not exclusively detect DSB repair, it does capture the incorporation of new nucleotides into the repaired DNA strand, which can occur with nonhomologous end joining[32]. We observed enrichment of SAR-seq signal at NPAS4-bound sites and higher levels of repair (SAR-seq signal)[32] at NPAS4-bound sites relative to FOS-bound sites (Extended Data Fig. 9f). To further investigate repair mechanisms at NPAS4-bound sites, we performed CUT&RUN for components of the MRE11–RAD50–NBS1 (MRN) complex. This complex is an early responder to sites of DSBs and plays an important part in processing broken DNA ends and initiating multiple repair pathways[33]. We examined both the MRE11 subunit, which mediates the removal of lethal topoisomerase cleavage products[34] and serves as a marker for DSB repair across the genome[35], and RAD50, which facilitates assembly of the complex and potentiates the endonuclease activity of MRE11 (ref. 33). After validating the specificity of the MRE11 CUT&RUN signal using *Mre11* cKO mice (Extended Data Fig. 10a,b), we observed a strong activity-dependent co-localization of MRE11 and RAD50 at NPAS4-bound elements that was not driven by nonspecific IgG binding, histone acetylation or chromatin accessibility (Fig. 4c,d and Extended Data Figs. 10c–f and 11a,b). Notably, MRE11 was also present at NPAS4-bound elements in the absence of stimulation, which suggests that it may be retained following stimulation or that these sites are primed with other repair factors. These findings from multiple independent assays indicate that NPAS4–NuA4-bound sites are hotspots of DNA damage and repair in activated neurons.

Fig. 4: NPAS4–NuA4-bound sites undergo DNA repair as inducible transcription subsides.



a, PCA of paired RNA-seq prepared from the same tissue used to generate the sBLISS-seq time course in hippocampus. Samples clustered according to the stimulation state, with 10 h stimulation samples separating into groups with either dampening or sustained transcriptional induction. DeSeq2-normalized counts of inducible genes *Bdnf* and *Nptx2* shown for each sample (0 h, $n = 8$; 2 h, $n = 8$; 10 h, $n = 10$). One mouse per replicate. **b**, Boxplots of average sBLISS-seq-normalized counts in 0 h ($n = 8$), 2 h ($n = 8$), 10 h less active ($n = 4$), and 10 h still active ($n = 6$) samples at activity-inducible regulatory elements, subset by quartiles of NPAS4 binding. One mouse per replicate. Boxplots show the median (line), IQR (box) and 1.5× IQR (whiskers), and notches indicate the median ± 1.58× IQR/sqrt(n). Q1 = 1,017 sites, Q4 = 764 sites. Q1, *$P = 0.03435$, **$P = 2.049 \times 10^{-8}$, ***$P = 3.057 \times 10^{-13}$; Q4, ***$P < 2.2 \times 10^{-16}$, not significant (NS) = 0.1742. P values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests.

c, IGV tracks of sBLISS-seq, γH2AX ChIP–seq, NPAS4, EP400, MRE11 and RAD50 CUT&RUN signals. Data plotted show the aggregate signals from all replicates per factor. The *y* axis displays normalized coverage scaled for each factor at the chosen locus. CUT&RUN and γH2AX ChIP–seq, n = 3–5 mice pooled per replicate. sBLISS-seq, $n = 1$ mouse per replicate. Replicate numbers provided in Supplementary Table 2. **d**, Boxplots of average MRE11 and RAD50 normalized counts at activity-inducible regulatory elements, subset by quartiles of NPAS4 CUT&RUN signals. Q1 = 1,017 sites, Q4 = 764 sites. Boxplots show the median (line), IQR (box) and 1.5× IQR (whiskers), and notches indicate the median ± 1.58× IQR/sqrt(n). n = 3–5 mice pooled per replicate. Replicate numbers provided in Supplementary Table 2. $^{***}P < 2.2 \times 10^{-16}$. P values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests.

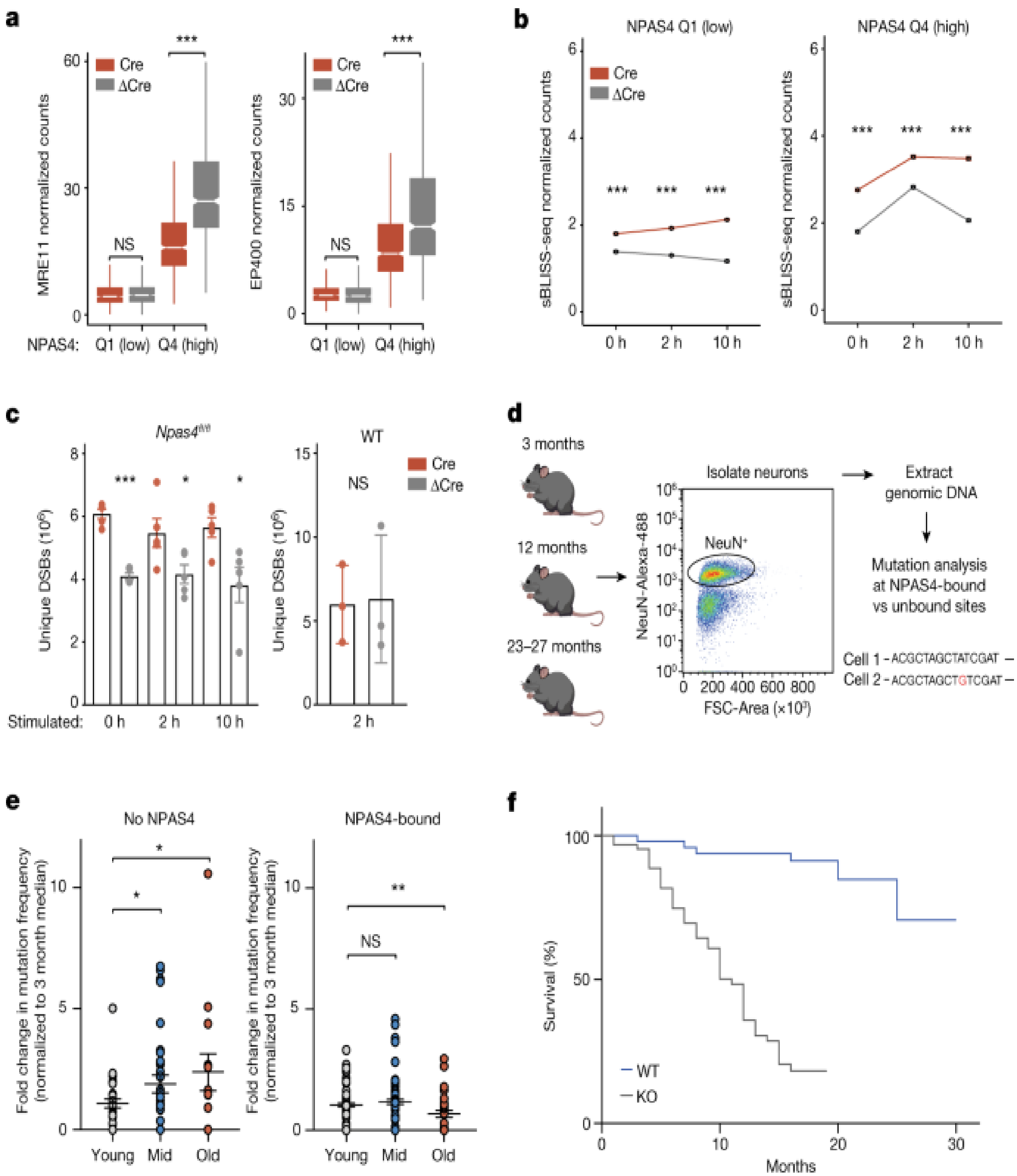
Source Data

NPAS4–NuA4 disruption impairs DSB repair

Given its targeting to sites of damage, we next asked whether NPAS4–NuA4 stimulates repair at these sites in part by recruiting additional DSB repair machinery to the genome. By performing CUT&RUN for the NuA4 components EP400 and MRE11, we observed that depletion of NPAS4 resulted in a significant reduction in the binding of both proteins to NPAS4–NuA4 sites (Fig. 5a and Extended Data Fig. 11c–h). Although MRE11 was significantly reduced, it was not completely abolished, which suggests that MRE11 may have additional targeting mechanisms, such as direct interaction with free DSB ends or association with transcriptional machinery[33,36]. If NPAS4–NuA4 stimulates repair, the depletion of NPAS4–NuA4 subunits should result in increased DSBs in neurons. To test this prediction, we injected *Npas4*$^{fl/fl}$ mice with either Cre-expressing or ΔCre-expressing AAVs, isolated nuclei at 0, 2 or 10 h after stimulation and performed sBLISS-seq. After short-term depletion of NPAS4, the number of DSBs at activity-inducible regulatory elements increased at 2 and 10 h after stimulation. This result suggests that loss of NPAS4 renders neurons unable to efficiently repair these transcription-coupled breaks (Fig. 5b). We also observed an increase in the number of breaks before stimulation, which may reflect accumulated effects of dysregulated repair during previous

stimulation (Fig. 5b and Extended Data Fig. 12a–d). The DSB increase in NPAS4-depleted nuclei at activity-inducible sites was not observed following Cre expression in wild-type neurons (Extended Data Fig. 12b). Of note, deletion of NPAS4 or the NuA4 component TIP60 also resulted in a significant increase in genome-wide DSBs that was not observed in wild-type mice and was not a secondary consequence of increased cell apoptosis in *Npas4* or *Tip60* cKO cells (Fig. 5c and Extended Data Fig. 12d–g). This increased level of DSBs across the genome may be due in part to dysregulated neuronal inhibition resulting from NPAS4–NuA4 disruption. However, we cannot rule out the possibility that expression of Cre, which has previously been reported to cause off-target DNA breaks on the genome[37], has a more pronounced effect in NPAS4 and TIP60 mutants owing to broadly dysregulated repair signalling.

Fig. 5: NPAS4–NuA4 disruption impairs genome stability and reduces lifespan.



a, Boxplots of average MRE11 or EP400 CUT&RUN normalized counts in Cre-infected or ΔCre-infected hippocampi of *Npas4*$^{fl/fl}$ mice at activity-inducible sites, subset by quartiles of NPAS4 binding. Q1 = 1,017 sites, Q4 = 764 sites. Boxplots show the median (line), IQR (box) and 1.5× IQR (whiskers), and notches indicate the median ± 1.58× IQR/sqrt(n). n = 2–3 mice pooled per replicate. Replicate numbers provided in Supplementary Table 2. MRE11 Q1, P = 0.2453; Q4, ***$P < 2.2 \times 10^{-16}$. EP400 Q1, P =

0.9962; Q4, ***$P < 2.2 \times 10^{-16}$. P values were calculated using unpaired, one-tailed Wilcoxon rank-sum tests (ΔCre > Cre). **b**, Line plot depicting average ± s.e.m. of sBLISS-seq normalized counts in Cre-infected or ΔCre-infected hippocampi of *Npas4*$^{fl/fl}$ mice, subset by quartiles of NPAS4 binding. Q1 = 1,017 sites, Q4 = 764 sites. Data from $n = 5$ each for Cre-infected and ΔCre-infected mice. ***$P < 2.2 \times 10^{-16}$. P values were calculated using unpaired, one-tailed Wilcoxon rank-sum tests (Cre > ΔCre). See Extended Data Fig. 12a for boxplot distribution of data. **c**, Genome-wide breaks in hippocampal nuclei isolated from *Npas4*$^{fl/fl}$ ($n = 5$) or wild-type ($n = 3$) mice infected with Cre or ΔCre virus. Data are mean ± s.e.m. *Npas4*$^{fl/fl}$: 0 h, $P = 5.24 \times 10^{-6}$; 2 h, $P = 0.044$; 10 h, $P = 0.022$. Wild-type: 2 h, $P = 0.906$. P values were calculated using two-tailed, unpaired *t*-tests. **d**, Collection of hippocampal neuronal nuclei across lifespan. **e**,Normalized mutation frequency at NPAS4-bound and unbound sites in wild-type mice. Mutation frequency (base changes and insertions and deletions) with age was calculated per site and normalized to the median frequency for that site in young animals. Points represent the normalized rate for one site sampled from one mouse. Data are mean ± s.e.m. Data are from $n = 4$ (young), 4 (middle aged) and 3 (old) mice. No NPAS4 sites: mid–young, *$P = 0.03$; old–young, *$P = 0.016$. NPAS4-bound sites: mid–young, NS = 0.18; old–young, **$P = 0.0095$. P values calculated using one-tailed, unpaired *t*-tests. **f**, Survival of *Npas4* wild-type ($n = 53$; 28 females, 25 males) and *Npas4* knockout ($n = 64$; 37 females, 27 males) mice, showing abbreviated lifespan of *Npas4* knockout mice. $P = 4.09 \times 10^{-13}$ by a two-tailed Mantel–Cox test, $P = 3.63 \times 10^{-11}$ by a two-tailed Gehan–Breslow–Wilcoxon test.

Source Data

Age-dependent mutations at NPAS4 sites

A distinctive feature of neurons is their long postmitotic lifespans, which provide ample time for the accumulation of unresolved DNA breaks and mutations. We wondered whether the repeated activation of NPAS4-bound elements could predispose these sites to increased mutational load during ageing. To assess the mutational load at NPAS4-bound elements, we used fluorescence-activated cell sorting (FACS) to isolate NeuN-expressing

nuclei from the hippocampus of young (3 months old), middle-aged (12 months old) and old (23–27 months old) mice and extracted neuronal DNA (Fig. 5d and Extended Data Fig. 13a,b). We then performed targeted amplicon sequencing for sensitive mutation detection[38] at NPAS4-bound elements compared with negative control elements (that is, sites with little to no NPAS4 binding, termed NPAS4-unbound). Incorporation of unique molecular identifiers (UMIs) during PCR amplification enabled the detection of mutations attributable to a single allele template while excluding base changes arising from sequencing errors[38] (Extended Data Fig. 13c). We first validated the ability of this technique to detect mutations introduced at CRISPR–Cas9-induced cut sites in neurons (Extended Data Fig. 13d).

We next defined a panel of fragile sites that undergo routine breaks or repair (assessed by levels of γH2AX and/or MRE11 binding) and are bound by NPAS4–NuA4. We compared this dataset with a set of elements that do not share these signatures of damage and are not bound by NPAS4-NuA4, but are matched for levels of inducible H3K27ac, inducible chromatin accessibility and AT/GC content (Extended Data Fig. 13e,f). To account for differences between age groups in the representation of target sites amplified in our assay, each of which may display a different baseline rate of mutational frequency, we normalized the mutation rate of each site to the median mutation frequency of that same site in the young animals. Notably, sites not bound by NPAS4 accumulated mutations with age, showing an approximately twofold increase by 12 months. By contrast, our panel of NPAS4-bound sites did not appear to increase in mutational frequency with age (Fig. 5e). These data suggest that NPAS4–NuA4-bound sites are relatively protected against the additional mutations that accrue over the course of organismal ageing. These differences were not attributable to an unusually high rate of mutations in unbound sites, as we observed a higher rate of both single nucleotide variant and insertion and deletion events at NPAS4-bound sites relative to non-bound sites in young animals (Extended Data Fig. 13g). These results further corroborate that NPAS4–NuA4 is targeting fragile sites and raise the possibility that the NPAS4–NuA4 complex may be required as an additional layer of protection for these recurrently broken sites.

NPAS4–NuA4 disruption decreases lifespan

Our data showed that disruption of NPAS4–NuA4 function leads to dysregulation of activity-dependent gene expression, increased DNA breaks at activity-regulated promoters and enhancers, impaired localization of protective repair machinery and defects in pyramidal neuron somatic inhibition. Therefore, we reasoned that disruption of the NPAS4–NuA4 complex would ultimately have widespread, long-term consequences as animals age. These changes could include deleterious effects on genome integrity, excitatory/inhibitory balance and organismal lifespan. Notably, we observed that loss of this neuronal factor substantially shortened the lifespan of both male and female mice, leading to a median lifespan of 12 and 11 months, respectively (Fig. 5f). This result corroborates results from an independent *Npas4* knockout line[39] and demonstrates a clear effect on longevity for mice of both sexes (Extended Data Fig. 13h,i). That the reduced lifespan of the germline *Npas4* knockout mice is due specifically to the loss of NPAS4 in the brain is buttressed by our snRNA-seq data demonstrating that NPAS4 is highly specific to neurons (Extended Data Fig. 2f). Moreover, mice with *Npas4* deleted in forebrain *Camk2a*-expressing excitatory neurons also had reduced lifespan (Extended Data Fig. 13j). However, we cannot exclude the possibility that transient expression of *Npas4* in non-neuronal cells contributes to this longevity phenotype. Together, these data raise the possibility that the protective role of NPAS4–NuA4 in facilitating DSB repair helps to ensure the long-term fidelity of transcriptional responses to stimulation and proper inhibitory control in the brain that may be crucial for normal lifespan.

Discussion

The extent to which different cell types in the body specialize their DNA repair mechanisms is poorly understood. Emerging evidence suggests that neurons continuously repair DNA at select locations within the genome[32,40], yet the mechanisms for preferential targeting of repair remained obscure. This may be due in part to the heterogeneity of recurrently damaged sites across the vast number of neuronal cell types resulting from their diverse activity patterns and cell-type-specific transcriptional programmes. Using a

combination of new mouse models, biochemistry, single-cell genomics and electrophysiology, we identified a specialized neuronal form of the NuA4 complex that assembles around NPAS4 in activated neurons to regulate cell-type-specific inducible transcription and suppress DNA damage. Our findings suggest that neurons have evolved a specific chromatin regulatory mechanism that couples synaptic activity to genome preservation. This mechanism may reduce accumulating damage at pivotal regulatory elements in each neuronal cell type and preserve the ability to mount appropriate responses to environmental cues.

The mechanisms that lead to both the formation and repair of DSBs at NPAS4-bound sites remain to be fully elucidated. As previously suggested[4], these breaks may arise from pre-bound topoisomerase enzymes that are post-translationally modified downstream of neuronal activity. In addition, DSBs may form in the process of resolving DNA–RNA hybrids (R-loops) or releasing stalled transcription complexes that occur with a rapid induction of previously quiescent regulatory elements. Notably, NuA4 has been reported to bind R-loops[41], which suggests that R-loop formation may contribute to both DNA damage and repair at activity-dependent regulatory elements. Studies outside the nervous system have suggested that RNA itself could facilitate repair of these transcribed regions by serving as a template in place of a sister chromatid in postmitotic cells[42,43]. Although it is probable that canonical nonhomologous end joining pathways mediate much of the DSB repair in activated neurons, it is possible that NPAS4–NuA4 engages multiple repair pathways. Future studies that probe the precise mechanisms that neurons use to repair activity-induced damage, including those mediated by NPAS4–NuA4, will be important areas of investigation.

This work provides an example of specialized chromatin machinery in the brain and adds to the repertoire of neuronal epigenomic features that are frequently dysregulated in both neurodevelopmental and neurodegenerative disorders. Several components of NPAS4–NuA4 (*Ep400*, *Trrap*, *Actl6b* and *Tip60*) are mutated in neurodevelopmental and autism spectrum disorders[44,45,46]. Our discovery of a link between neuronal activity and DNA repair mediated by NPAS4–NuA4 suggests that damage at activity-dependent regulatory elements may be a source of neuronal dysfunction in these disorders.

Loss of genome integrity is a hallmark of ageing across organisms[12,13], and the ability to efficiently repair DSBs has been linked to the evolution of longer lifespans in mammals[47]. However, much remains unknown about how neuronal activity influences genome stability with age and contributes to cellular and organismal longevity. Sustaining neuronal vitality over time appears to require careful balancing of the proper ratio of excitation and inhibition[48]. In addition to the role of NPAS4 in DNA repair, the compounding loss of inhibition in *Npas4* knockout mice might contribute to their shortened lifespan. Over time, impaired NPAS4–NuA4 signalling may disrupt a key regulatory feedback loop in which deletion of NPAS4 impairs the expression of genes that mediate recruitment of somatic inhibition, which in turn leads to excessive excitation that further threatens genome stability. Future experiments that decouple the roles of NPAS4 in transcription and repair are needed to understand the relative contribution of these two processes in neuronal and organismal longevity. Given that the neuronal-specific expression pattern of NPAS4–NuA4 is conserved in the human brain, in which neurons are subject to recurrent activity-induced DNA breaks over many decades, the NPAS4–NuA4 signalling axis may also serve as an important entry point to understanding the breakdown of cognitive and sensory processing in ageing and neurodegenerative diseases in humans.

Methods

Experimental models

Mouse models

Animal use was approved and overseen by the Harvard University Institutional Animal Care and Use Committee and the Harvard Center for Comparative Medicine. The following mouse lines were used: wild-type C57/BL6 (Jackson Laboratory, 000664); *Npas4*$^{fl/fl}$[14]; *Npas4*-knockout[14]; *Tip60*$^{fl/fl}$ [49]; *Npas4–Flag-HA* (this manuscript); *Arnt2–Flag-HA* (this manuscript), *Tip60-H3F*[50]; *Mre11*$^{fl/fl}$[51,52]; *B6;129-Gt(ROSA)26Sor<tm5(CAG-Sun1/sfGFP)Nat>/J* (Jackson Laboratory, 021039)[53]; and *B6.Cg-Tg(Camk2a-cre)T29-1Stl/J* (Jackson Laboratory,

005329)[54]. Mice were housed in a temperature and humidity-controlled environment using ventilated microisolator cages. Mice were kept under a standard 12 h light–dark cycle, with food and water provided ad libitum. Male and female littermate mice were used in similar proportions and divided between control and experimental groups for all experiments conducted. No statistical methods were used to predetermine sample sizes. For biochemistry and genomic experiments, animals were collected at 4–6 weeks of age throughout the manuscript. For physiology experiments, animals were dissected and patched at postnatal day 24 (P24) to P28. For ageing experiments, animals were collected at 3-4 months, 12 months and 23–27 months of age. Details of animal age and sex are detailed within each protocol.

Generation of *Npas4–Flag-HA* and *Arnt2–Flag-HA* knock-in mouse lines

Zygote injections were performed at the Harvard Genome Modification Facility in accordance with their practices and guidelines. Guide RNA sequences in proximity to the 3′ end of *Npas4* and *Arnt2* loci were chosen based on predicted cutting efficiency and low off-target effects. Guide RNA sequences for NPAS4 (5′-cacagacttattcaaaacgt-3′) and ARNT2 (5′-gagtagcttcaggcaaagcc-3′) were cloned into a FUGW backbone containing the guide sequence upstream of a tracrRNA (5′-gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtggtgcttttttt-3′). To generate templates for in vitro transcription, PCR was performed using these FUGW plasmids to incorporate a T7 promoter using the following primers: *Npas4* forward (F): 5′-taatacgactcactataggcacagacttattcaaaacgtgttttagag-3′; *Npas4* reverse (R): 5′-aaaaaaagcaccaccgactcgg-3′; *Arnt2* F: 5′-taatacgactcactataggagtagcttcaggcaaagcc-3′; and *Arnt2* R: 5′-aaaaaaagcaccaccgactcgg-3′.

PCR products were verified on DNA gels to ensure a single product and were purified using Qiagen's PCR clean-up kit. In vitro transcription (IVT) was performed using 400 ng of purified RNA template for 16 h at 37 °C using a Megascript Short Transcript IVT kit (Thermo Fisher, AM1354). RNA clean-up was performed using a MEGAclear Transcription Clean-Up

kit (Ambion, AM190). Finally, RNA size and quality were assessed by running 200 ng of purified product on 10% TBE gels. The HDR template was ordered as a 200 bp, single-stranded Ultramer from IDT, with 75 bp of homology on either side of the cut site. CRISPR guide RNA, Cas9 RNA and Ultramer donor template were injected into zygotes of mixed background DBA/C57BL6 mice according to the guidelines and procedures of the Harvard Genome Modification facility. DNA from founders and F_1 progeny were screened using Sanger sequencing to ensure proper insertion of the tag and that no mutations occurred in the flanking DNA regions. *Npas4–Flag-HA* and *Arnt2–Flag-HA* mice have been subsequently backcrossed to C57BL6/J mice and are available upon request.

Mouse neuron culture

Embryonic cortices (embryonic day 16.5–17) were dissected from 5–6 embryos of mixed sex and pooled for a single replicate culture. Papain (Sigma-Aldrich, 10108014001) was used to dissociate tissue through a 10 min incubation at 37 °C. Digestion was terminated with the addition of ovomucoid (trypsin inhibitor, Worthington). Digested tissue was gently triturated through a P1000 pipette to release cells and passed through a 40 µm filter. Neurons were plated onto cell culture dishes pre-coated overnight with poly-d-lysine (20 µg ml^{-1}) and laminin (4 µg ml^{-1}). Neurons were grown in neurobasal medium (Gibco) containing B27 supplement (2%), penicillin–streptomycin (50 U ml^{-1} penicillin and 50 U ml^{-1} streptomycin) and glutaMAX (1 mM). Neurons were grown in incubators maintained at 37 °C with a CO_2 concentration of 5%. Neurons were collected for all experiments at 7 days in vitro (DIV), and fresh medium was added at 3–4 DIV (50% total volume) unless otherwise indicated. Independent replicates were generated by preparing primary cultures from pools of embryos dissected on different days and maintained in culture for 7 DIV. For RNA-seq experiments, cells were plated at a density of 156,250 cells per cm^2 (1.5 million neurons per well) in 6-well culture dishes. For luciferase assays, cells were plated at a density of 210,526 cm^{-2} (400,000 cells per well) in 24-well culture dishes. To depolarize neurons, neurons were treated with 55 mM KCl. For treatment with different stimuli (Extended Data Fig. 1d), 7 DIV neurons were treated with 50 µM NMDA or 50 ng ml^{-1} BDNF. To

chelate calcium, neurons were pretreated with 5 mM EGTA 15 min before stimulation with 55 mM KCl solution.

Experimental procedures

Collection time points

For experiments in which seizures were induced, KA (5–8 mg kg^{-1} (electrophysiology) or 12–15 mg kg^{-1} (gene expression, CUT&RUN, sBLISS-seq and immunoprecipitations) (Sigma-Aldrich, K0250) was intraperitoneally injected. For genomic-level analyses, mice were euthanized 2, 6 or 10 h after injection, as indicated, based on previous work showing that immediate-early gene transcription factors and late-response gene programmes are induced at 2 and 6 h after stimulus onset, respectively[1,14,28]. The unstimulated or basal time point indicates mice injected with saline control and were dissected and processed in parallel with the stimulated tissue. For electrophysiological experiments in CA1 pyramidal neurons, mice were killed 18–24 h after injection with low levels of KA to allow sufficient time for the expression of NPAS4, its associated activity-dependent target genes and the execution of potential synaptic regulation. For proteomics analysis, mice were killed 3 h after injection to allow sufficient time for NPAS4 to be produced and for the assembly of its associated protein complex.

Stereotaxically guided surgery

For genomic assays, including snRNA-seq, CUT&RUN and sBLISS-seq, P21–P24 *Tip60*$^{fl/fl}$ and *Npas4*$^{fl/fl}$ mice were anaesthetized by isoflurane inhalation (3–5% for induction, 1–3% maintenance) and positioned within a stereotaxic frame (Kopf) in which the temperature of the mouse was maintained at around 37 °C with a heat pad. All surgeries were performed according to protocols approved by the Harvard University Standing Committee on Animal Care and were in accordance with federal guidelines. Fur around the scalp was removed using a shaver and sterilized with three alternating washes of betadine and 70% ethanol. For genomic assays requiring broad infection of the hippocampus, a burr hole was drilled

through the skull above the hippocampus, and a glass pipette filled with AAV was lowered to a central region of the hippocampus to enable broad infection of the various subregions (medial–lateral: ±2.5; anterior–posterior: −2.5; dorsal–ventral: −2.5). AAV (1,000 nl, diluted to 1.0×10^{12} genome copies per ml) was injected at 150 nl min^{-1}, and the pipette was left in place for 5 min after completion of virus infusion to allow for viral spreading. All animals were given postoperative analgesic (buprenophine slow-release formulation, 1 mg kg^{-1}) in accordance with Institutional Animal Care and Use Committee protocols. For electrophysiological assays requiring sparse infection of the CA1 region of the hippocampus, P14 mice were injected as described above, but the glass needle tip was positioned at different coordinates to limit infection to the CA1 (medial–lateral: ±3.4; anterior–posterior: −2.9; dorsal–ventral: −2.8), and a lower titre of virus was injected (1.0×10^{10} genome copies per ml).

Acute slice preparation

Transverse hippocampal slices were prepared from *Npas4*$^{fl/fl}$ and *Tip60*$^{fl/fl}$ C57BL/6 mice aged P21–P28. Isoflurane was used to anaesthetize the animals, and the brain was quickly removed, bisected along the midline and placed in ice-cold choline-containing artificial cerebrospinal fluid (choline-ACSF) containing (in mM): 110 choline-Cl, 25 $NaHCO_3$, 1.25 NaH_2PO_4, 2.5 KCl, 7 $MgCl_2$, 25 glucose, 0.5 $CaCl_2$, 11.6 ascorbic acid and 3.1 pyruvic acid. Choline-ACSF was equilibrated with 95% O_2 and 5% CO_2. Both cerebral hemispheres were transferred to a slicing chamber containing ice-cold choline-ACSF, and 300-µm-thick slices were cut using a Leica VT1200S vibratome. Slices were transferred to a holding chamber filled with ACSF saturated with 95% O_2 and 5% CO_2 and containing (in mM): 127 NaCl, 25 $NaHCO_3$, 1.25 NaH_2PO_4, 2.5 KCl, 2 $CaCl_2$, 1 $MgCl_2$ and 25 glucose. Slices were incubated at 34 °C for 30 min, then equilibrated to room temperature for 20 min before recordings. All recordings were performed at room temperature and were completed within 6 h of slice preparation. Epifluorescence was used to identify slices with sparse infection of the CA1 pyramidal layer with Cre-mCherry, and slices with >30% infected CA1 pyramidal neurons were not used for recordings. Slices were also discarded

if Cre-mCherry expression was observed in the CA3 or dentate gyrus regions.

Electrophysiology

CA1 pyramidal neurons were visualized with infrared differential interference contrast microscopy to perform whole-cell voltage-clamp recordings. Neighbouring uninfected and Cre-mCherry-infected neurons were identified using epifluorescence driven by a light-emitting diode. Neurons were held at −70 mV for all experiments. Recording pipettes made from borosilicate glass (open resistance between 2 and 4 MΩ) were filled with an internal solution containing (in mM): 147 CsCl, 5 Na_2-phosphocreatine, 10 HEPES, 2 MgATP, 0.3 Na_2GTP, 2 EGTA and 5 QX-314 (Sigma-Aldrich). Internal solution was prepared in a single batch, with osmolarity adjusted to 290 mOsm with double-distilled water and pH adjusted to 7.3 with CsOH, and was stored at −20 °C. To record IPSCs, inhibitory currents were pharmacologically isolated through bath application of 10 μM (*R*)-CPP (Tocris Bioscience) and 10 μM NBQX disodium salt (Tocris Bioscience). Extracellular stimulation of perisomatic inhibitory axons was achieved using a concentric bipolar electrode (FHC) placed within the centre of the stratum pyramidale and within 100–200 μm laterally of the pair of voltage-clamped cells. The stimulus strength used was the minimum stimulation required to generate a reliable IPSC in both neurons.

Electrophysiology data acquisition and analysis

Recordings were made using a Multiclamp 700B amplifier (Axon Instruments), filtered at 3 kHz and sampled at 10 kHz. Data analysis was performed with custom software written in MatLab. Experiments were discarded if the holding current was less than −600 pA or if the series resistance was greater than 25 MΩ. Series resistance across simultaneously recorded cells was within 25% for each pair. Recordings were performed at room temperature (19–21 °C). The amplitude of IPSCs was calculated by averaging the amplitude 0.5 ms before to 2 ms after the peak of the current. The unsigned magnitude of synaptic currents are shown for clarity.

Histology

Males and females were used in equal proportions. Mice were anaesthetized by intraperitoneal injection of 10 mg ml^{-1} ketamine and 1 mg ml^{-1} xylazine. Once anaesthetized, mice were transcardially perfused with at least 10 ml of ice-cold PBS, followed by at least 20 ml of 4% paraformaldehyde (PFA) in PBS. Brains were removed and placed in 4% PFA at 4 °C for 24 h, followed by three successive washes in cold PBS. Brains were then transferred to 30% sucrose for cryoprotection, and were incubated at 4 °C until the tissue equilibrated and sank. The brains were then embedded in NEG-50 frozen section medium and stored at −80 °C until sectioned (30 μm thick) on a cryostat (Leica CM1950). Sections were stored in PBS at 4 °C until further use. For immunohistochemistry, sections were permeabilized and blocked by incubating in PBS containing 0.1% Triton X-100 and 5% normal goat serum (blocking solution) for 1 h at room temperature. Staining in primary antibody (dilutions listed below) was performed overnight (16 h) at 4 °C with gentle shaking. Sections were then washed three times for 5 min each in PBS containing 0.1% Triton X-100. Secondary antibody staining was performed by diluting an Alexa Fluor dye-conjugated secondary antibody (Life Technologies; rat Alexa Fluor 555 (A21434), rabbit Alexa Fluor 647 (A31573); 1:250) and incubating slices for 2 h at room temperature. Sections were then washed again three times for 5 min each in PBS containing 0.1% Triton X-100, and then mounted in DAPI Fluoromount-G (Southern Biotech) and imaged on a slide-scanning microscope (Olympus VS120, VS ASW-FL). Antibodies were diluted in blocking solution as follows: rat anti-HA (Sigma-Aldrich, ROAHAHA; 1:250); rabbit anti-NPAS4 (in house; 1:1,000)[14]; rabbit anti-ARNT2 (in house; 1:1,000)[28]; rabbit anti-KAT5 (TIP60) (Proteintech, 10827-1-AP; 1:250); rabbit anti-cleaved caspase-3 (Cell Signaling Technology, 9664S; 1:1,000).

Immunoprecipitation of NPAS4–NuA4 complexes

To isolate protein complexes associated with NPAS4, 8–12 hippocampi were isolated from 6-week-old *Npas4–Flag-HA* and *Arnt2–Flag-HA* mice injected with 15 mg kg^{-1} KA (Sigma-Aldrich, K0250) to induce high levels of NPAS4 expression. Males and females were used in equal proportions in pooled replicate samples. Wild-type mice of the same age and sex

distributions were processed in parallel to serve as controls. Hippocampal regions were dissected 3 h after KA injection to allow sufficient time for NPAS4 expression and assembly into potential complexes. Hippocampi were collected and dounced 20× in 10 ml of NE1 buffer (20 mM HEPES pH 7.9, 10 mM KCl, 3 mM $MgCl_2$, 0.1% Triton and 0.1 mM EDTA) containing protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktails 2 and 3 (Sigma-Aldrich, P5726 and P0044). Nuclei were released through 10 min of incubation in NE1 buffer and then pelleted by gentle centrifugation (1,000–1,500*g*). Nuclear pellets were resuspended in 1 packed nuclear volume of NE1 buffer. To facilitate release of chromatin-associated proteins, nuclei were incubated for 30 min at 4 °C with benzonase endonuclease (Sigma-Aldrich, E1014, >25 KU (1 µl per 10 million nuclei)) followed by the addition of NaCl to a final concentration of 300 mM. Following high-speed centrifugation to remove insoluble material, lysates were diluted to achieve a final salt concentration of 200 mM. To preclear nonspecific interactions, lysates were incubated for 30 min at 4 °C with 50 µl of ms-IgG-coated agarose beads (Sigma-Aldrich, A0919). Following preclear, lysates were incubated for 1.5 h with 60 µl of anti-M2-Flag resin (Sigma-Aldrich, A2220) per 1 ml of diluted lysate. Samples were washed 4× in NE1 buffer containing 250 mM NaCl for 5 min with rotation at 4 °C. NPAS4-interacting or ARNT2-interacting proteins were competitively eluted off M2 resin by incubation in 50–100 µl of 500 µg ml^{-1} 3× Flag peptide (Sigma-Aldrich, F4799) diluted in NE1 buffer for 30 min at room temperature. For mass spectrometry analysis, eluted proteins were precipitated with trichloroacetic acid. Replicates shown in Fig. 1c consisted of 3 independent pools of 8–12 mice collected from wild-type or *Npas4–FH* lines and processed on separate days. Validation immunoprecipitation assays using either *Npas4–Flag-HA* mice or *Tip60-H3F* mice followed by immunoblotting analysis were conducted under the same conditions as described above. For validation experiments in mouse visual cortex following light stimulation, 20 hippocampi were isolated from the V1 cortices of *Npas4–Flag-HA* or wild-type controls. Mice were housed in the dark for 1 week followed by 2 h of light stimulation. Data are shown in Extended Data Fig. 1f.

To isolate high molecular weight complexes containing intact NPAS4–NuA4 for mass spectrometry, 8–10 6-week-old *Npas4–Flag-HA*, *Tip60-H3F* and wild-type control mice were injected with KA and dissected 2 h after

injection. Males and females were used in equal proportions in pooled replicate samples. Replicates consisted of independent pools of 8–10 mice run on independent gradients and days. Two replicates were performed. Whole forebrain, minus hippocampus and striatum, were minced and transferred NE1 buffer, and nuclear lysates were prepared as described above. Approximately 6 ml gradients were prepared using a BIOCOMP Gradient Master 108 (Science Services) using a 10–40% long glycerol gradient for the SW41 Beckman rotor. Sample concentrations were normalized to around 2 mg ml^{-1}, and 1 ml was loaded on top of the 10–40% glycerol gradient. Samples were centrifuged for 16 h at 37,000 r.p.m. at 4 °C. One millilitre fractions were collected at 4 °C. Fractions were run on SDS–PAGE gels to assess which fractions contained components of the NuA4 complex. Immunoprecipitation assays were then performed from fractions 1–3, which contained the majority of TRRAP and EP400 in the lysates. To preclear nonspecific interactions from fractions, lysates were incubated for 30 min at 4 °C with 100 µl of ms-IgG-coated agarose beads (Sigma-Aldrich, A0919). Following preclear, lysates were incubated for 1.5 h with 100 µl of anti-M2-Flag resin (Sigma-Aldrich, A2220) per 1 ml fraction. Samples were washed 4× in NE1 containing 250 mM NaCl for 5 min with rotation at 4 °C. Interacting proteins were eluted off M2 resin by incubation in 500 µg ml^{-1} 3× Flag peptide (Sigma-Aldrich, F4799). Eluates from immunoprecipitation assays performed on fractions 1–3 were pooled to have sufficient material for mass spectrometry, and eluted proteins were precipitated with trichloroacetic acid.

Immunoprecipitation of NPAS4 truncation mutants in HEK293T cells

Plasmids (FUW with UbC promoter) containing the sequence for Flag-HA-tagged FOS, JUN, EGR1 and NPAS4 or various truncated forms of NPAS4 (Supplementary Table 1) were expressed in HEK293T cells by $CaCl_2$–BBS transfection (2 µg DNA, 6×10^6 cells). HEK293T cells were obtained from Thermo Fisher Scientific (50188404FP). HEK293T cells were not authenticated or tested for mycoplasma. With the exception of nuclear GFP-FH control samples, the plasmids expressed GFP–IRES-NPAS4 to visualize transfection efficiency. Flag-HA tags were appended to the amino terminus of NPAS4 to minimize differences in tag accessibility across NPAS4 variants truncated at the C-terminal end. Cells were collected by gently

scraping into ice-cold PBS containing protease inhibitor cocktail (Roche) and were pelleted by gentle centrifugation (2,000*g*). Nuclei were isolated, and Flag immunoprecipitation was performed as described above. A second immunoprecipitation step using HA was performed by collecting the Flag peptide eluate, increasing the volume to 1 ml in NE1 buffer, adding 50 µl anti-HA resin (Santa Cruz Biotechnology, SC-7392 AC) and gently rotating at 4 °C for 1.5 h. Samples were washed 4× in NE1 buffer containing 250 mM NaCl for 5 min with rotation at 4 °C. Proteins were eluted by incubating resin with HA peptide (Thermo Fisher, 26184) diluted in NE1 buffer for 30 min at room temperature. Eluted proteins were precipitated with trichloroacetic acid for mass spectrometry. Replicates consisted of independently transfected cultures and immunoprecipitation assays performed on separate days.

Size-exclusion chromatography and blue native gels

Whole-cell lysates (2% Triton X-100, 50 mM Tris, 150 mM NaCl and 1 mM EDTA) from hippocampal tissue or from hippocampal and striatal tissue were fractionated using a 400HR 1 fractionator set to 0.5 ml min^{-1}. Buffers for the column and fractionator consisted of 20 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EDTA, 3 mM $MgCl_2$ and 0.02% Triton. Seventeen fractions consisting of 6 ml each were collected, and fractions 8–16 were loaded onto Native Novex NuPAGE 4–16% gels (Thermo Fisher, BN1002BOX) to estimate the approximate size of NPAS4-containing protein complexes. Native page marker (1 µl; Invitrogen, LC0725) was used to estimate sizes of the complexes. Proteins were transferred onto PVDF membranes and immunoblotted for NPAS4 (1:1,000) using in-house antibodies.

Immunoblotting

Whole-cell or nuclear extracts from primary neurons or brain tissue were resolved on 3–8% Tris-acetate gels (Thermo Fisher, EA0375BOX) or 10% Tris-glycine gels and transferred to nitrocellulose membranes. Membranes were incubated overnight in the following primary antibodies: EP400 (Bethyl Labs, A300-541-A; Abcam, Ab5201; Abcam, 70301; 1:1,000); DMAP1 (Cell Signaling Technology, 13326; Santa Cruz, sc-373949; 1:500

or 1:1,000); TRRAP (Bethyl, A301-132A; 1:500 or 1:1,000); ARNT2 (in house; 1:1,000)[28]; NPAS4 (in house, 1:1,000)[14]; FOS (in house; 1:1,000); β-tubulin3 (Covance, MMS-435P; 1:5,000); GAPDH (Sigma-Aldrich, G9545; 1:5,000); histone H3 (Abcam, 1791; 1:10,000); NPAS3 (gift from S. McKnight; 1:1,000)[55]; and HA (Cell Signaling Technology, C29F4; 1:1,000). Following washing, membranes were incubated with secondary antibodies conjugated to IRdye 700 or 800 and imaged using a LiCOR Odyssey instrument.

Mass spectrometry

Samples were processed according to standard procedures of the Taplin Mass Spectrometry Facility (Harvard University). Rehydrated proteins were incubated with 50 mM ammonium bicarbonate solution containing 12.5 ng μl^{-1} modified sequencing-grade trypsin (Promega) at 4 °C for 45 min. Following removal of excess trypsin solution, samples were placed at 37 °C in 50 mM ammonium bicarbonate solution. Peptides were recovered through the removal of ammonium bicarbonate solution and were subsequently washed in 50% acetonitrile and 1% formic acid before dehydration.

Lentivirus production

To generate lentiviral shRNA expression constructs, 21 bp targeted sequences from the TRC (Sigma-Aldrich) were subcloned into the FUW vector downstream of the U6 promoter. The following sequences were used: control/non-targeting shRNA (5′-gcgcgatagcgctaataattt-3′); *Npas4* shRNA-1 (5′-ggttgaccctgataatttta-3′); *Tip60* shRNA-1 TRCN0000039299 (5′-cctcctatcctaccgaagtta-3′); *Tip60* shRNA-2 TRCN0000039300 (5′-cggagtatgactgcaaaggtt-3′); *Ep400* shRNA-1 TRCN0000109315 (5′-ccgtgaacattagctttgatt-3′); and *Ep400* shRNA-2 TRCN0000305480 (5′-gtcgtcagaaggccttatatg-3′).

To generate LentiCRISPR constructs used for testing sBLISS-seq in cultured neurons, the following guide sequences were cloned into LentiCRISPRv2GFP (Addgene, 82416) using the listed primer sequences: (1) *Scg2* 5′-cggcccgagccctcactca-3′: primer F: 5′-caccgcggcccgagccctcactcag-3′; primer R: 5′-aaacctgagtgagggctcgggccgc-3′;

(2) *Inhba* enhancer 5′-gagcagccactagcgaaccc-3′: primer F: 5′-caccggagcagccactagcgaaccc-3′; primer R: 5′-aaacgggttcgctagtggctgctcc-3′; (3) *Bdnf* 5′-tgatagtggaaattgcatg-3′: primer F: 5′-caccgtgatagtggaaattgcatgg-3′; primer R: 5′-aaacccatgcaatttccactatcac-3′; (4) *Fos* 5′-gcgcggtcactgctcgttc-3′: primer F: 5′-caccggcgcggtcactgctcgttcg-3′; primer R: 5′-aaaccgaacgagcagtgaccgcgc-3′.

To produce lentivirus for shRNA-mediated depletion of NPAS4, EP400 and TIP60 in primary neuronal cultures or to express LentiCRISPR constructs, 10 µg of lentiviral plasmid was transfected into HEK293T cells along with third-generation packaging plasmids pMDL (5 µg), RSV (2.5 µg) and VSVG (2.5 µg). For the LentiCRISPR pool, 2.5 µg each of LentiCRISPR *Fos* gRNA, LentiCRISPR *Scg2* gRNA, LentiCRISPR *Bdnf* gRNA and LentiCRISPR *Nptx2* gRNA were transfected, for a total of 10 µg of plasmid. At 12–16 h following transfection, transfected medium was exchanged for fresh medium, and supernatant containing virus was collected at 48 h after transfection. For shRNA constructs, supernatant containing virus particles from 10–15 plates of transfected HEK293T cells were pooled, and virus particles were isolated by high-speed centrifugation (25,000 r.p.m. for 90 min). Pelleted virus was resuspended overnight at 4 °C in 100–150 µl of 1× PBS. For LentiCRISPR constructs, supernatant containing virus particles from 10 plates of transfected HEK293T cells were collected. Particles were precipitated by the addition of 1 volume of 4× PEG solution (40% PEG-8000, 1.2 M NaCl in 1× PBS pH 7.4) to 3 volumes of viral supernatant and stored for 1 h at 4 °C. Following incubation of the PEG and viral mixture, particles were precipitated by centrifugation at 1,500*g* for 45 min. Pelleted virus was resuspended overnight at 4 °C in 500 µl of 1× PBS. Individual viruses were titred, and the minimum amount of virus required to achieve approximately 85–100% infection, as assessed by GFP florescence, was determined for each lentivirus (1–10 µl per 1.5 million neurons for shRNA constructs, and 60 µl per 1.5 million neurons for LentiCRISPR constructs). Neurons were infected on 3 DIV with shRNA viruses and collected at 7 DIV for RNA-seq analysis. Neurons were infected on 2 DIV with LentiCRISPR viruses and collected at 7 DIV for sBLISS-seq analysis.

All AAV backbones were generated by using standard cloning and molecular biology techniques. AAV2/9 was prepared at the Boston Children's Hospital Viral Core.

Neuronal transfection and luciferase reporter assays

Luciferase induction was regulated by positioning NPAS4 target enhancers upstream of the luciferase gene. Sequences to test in luciferase reporter assays were chosen on the basis of the strength of NPAS4 binding in cultured neurons as previously described[28]. High-affinity NPAS4 sites include regions selected from the top 100 high-confidence NPAS4-binding peaks. Regions were PCR-amplified from mouse genomic DNA and subcloned into the pGL4.11 vector using standard Gibson assembly. See below for primer sequences.

To conduct the luciferase assays, 400,000 mouse hippocampal neurons were plated onto 24-well culture dishes and transfected with plasmids using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. At 5 DIV, neurons were transfected with 1 µg of total DNA consisting of 450 ng of firefly luciferase reporter DNA in pGL4.11, 50 ng pGL4.74 renilla luciferase reporter DNA (Promega) and 500 ng of NPAS4 overexpression construct. Lipofectamine (2 µl, Invitrogen) was used for each 1 µg of DNA. DNA–lipofectamine complexes were added dropwise to neurons and incubated for 2 h, after which the transfection medium was replaced with conditioned neuronal medium. Neurons were silenced on 6 DIV overnight through the addition of 1 µM TTX (Abcam, ab120055) and 100 µM AP5 (Thermo Fisher, 01-061-0). At 7 DIV, neurons were collected. In brief, neurons were washed 2× with PBS and lysed through the addition of 500 µl of passive lysis buffer (Dual-Luciferase Reporter Assay System, Promega, E1910). Next, 20 µl of each lysate was added to one well of a Costar white polystyrene 96-well assay plate (Corning). Luciferase Assay reagent II (LARII) and Stop & Glo reagent (Dual-luciferase assay system, Promega) were added to neuronal lysates, and luciferase/renilla measurements were made with a Synergy 4 Hybrid Microplate Reader (BioTek).

To control for variations in transfection efficiency and cell lysate generation, luciferase/renilla ratios were first calculated for each independently transfected well. Luciferase activity (luciferase/renilla) was then normalized to the average value of luciferase activity in nuclear GFP-expressing control samples collected on the same day from the same culture. Data were collected from at least three independent primary neuronal cultures

generated on separate days, except for peak 1 Npas4_1–699 (collected from two independent experiments). For each experimental culture, two to four independent wells were transfected. Data from each independently transfected well across all experiments are displayed. Error bars are ±s.e.m. *P* values were calculated using two-tailed, unpaired *t*-tests using in Prism (v.8.4.2). Benjamini–Hochberg correction for multiple hypothesis testing was performed in R (v.3.6.1).

Primers used the clone the listed genomic locations into pGL4.11 are listed below:

Peak 1: chromosome 5: 103,753,620–103,754,119

F: 5′-ggaaacagtgtctcacacaacctgcagggtttctgggcactttcaaaact-3′

R: 5′-caatgccaactgttactagctggcgcgcccgggcaccgggggcagccgggcgag-3′

Peak 2: chromosome 7: 112,679,812–112,680,311

F: 5′-ggaaacagtgtctcacacaacctgcaggagcggccccgccgccct-3′

R: 5′-caatgccaactgttactagctggcgcgcctttttggggtcctcctggg-3′

Peak 3: chromosome 8: 84,197,485–84,197,984

F: 5′-ggaaacagtgtctcacacaacctgcaggtgcctgctcgcttgttttacctcc-3′

R: 5′-caatgccaactgttactagctggcgcgcctaccagggcgctgttggggtgggtg-3′

Peak 4: chromosome 1:118,825,347–118,825,846

F: 5’-ggaaacagtgtctcacacaacctgcagggcaggttgatttggtaggaa-3’

R: 5′-caatgccaactgttactagctggcgcgccgtgcaaggagagccgcgggcg-3′

Peak 5: chromosome 2: 94,243,619–94,244,238

F: 5′-ggaaacagtgtctcacacaacctgcagggcgggtgtgcaacgggcagagag-3′

R: 5′-caatgccaactgttactagctggcgcgcctcatgtcccacccctaggcttcctc-3′

Peak 6: chromosome 12: 104,415,550–104,416,049

F: 5′-agtgtctcacacaacctgcaggcactgcagccagcgtccaccatgctt-3′

R: 5′-aactgttactagctggcgcgcccgactccatctttgacaaaag-3′

Validation by quantitative PCR with reverse transcription

RNA was extracted using a Qiagen RNeasy Micro kit (Qiagen, 74004), and equivalent amounts of RNA (100–200 ng) across all samples were converted to cDNA using a High Capacity cDNA kit (Thermo Fisher, 4368813) according to manufacturer's instructions. cDNA was diluted by at least threefold before running standard quantitative PCR (qPCR) with reverse transcription methods using Sybr Green master mix. qPCR was performed with technical triplicates using a QuantStudio 3 qPCR machine (Applied Biosystems). Expression for each qPCR target gene was normalized to the housekeeping genes *Gapdh* or *Tubb3* as indicated. The following primer sets were used to amplify genes of interest.

Primer sets:

Gapdh F: 5′-AGGTCGGTGTGAACGGATTTG-3′

Gapdh R: 5′-GGGGTCGTTGATGGCAACA-3′

Tubb3 F: 5′-TAGACCCCAGCGGCAACTAT-3′

Tubb3 R: 5′-GTTCCAGGTTCCAAGTCCACC-3′

S100b F: 5′-TGGTTGCCCTCATTGATGTCT-3′

S100b R: 5′-CCCATCCCCATCTTCGTCC-3′

Mog F: 5′-ACCTCTACCGAAATGGCAAGG-3′

Mog R: 5′-TCACGTTCTGAATCCTAAGGGT-3′

Aldh1l1 F: 5′-CAGGAGGTTTACTGCCAGCTA-3′

Aldh1l1 R: 5′-CACGTTGAGTTCTGCACCCA-3′

Pdgfra F: 5′-AGAGTTACACGTTTGAGCTGTC-3′

Pdgfra R: 5′-GTCCCTCCACGGTACTCCT-3′

Gfap F: 5′-CGGAGACGCATCACCTCTG-3′

Gfap R: 5′-AGGGAGTGGAGGAGTCATTCG-3′

Grin1 F: 5′-AGAGCCCGACCCTAAAAAGAA-3′

Grin1 R: 5′-CCCTCCTCCCTCTCAATAGC-3′

Rbfox3 F: 5′-ATCGTAGAGGGACGGAAAATTGA-3′

Rbfox3 R: 5′-GTTCCCAGGCTTCTTATTGGTC-3′

Grin2b F: 5′-GCCATGAACGAGACTGACCC-3′

Grin2b R: 5′-GCTTCCTGGTCCGTGTCATC-3′

Synapsin1 F: 5′-AGCTCAACAAATCCCAGTCTCT-3′

Synapsin1 R: 5′-CGGATGGTCTCAGCTTTCAC-3′

Npas4 F: 5′-ACCTAGCCCTACTGGACGTT-3′

Npas4 R: 5′-CGGGGTGTAGCAGTCCATAC-3′

Tip60 F: 5′-GTCACCCGGATGAAGAACAT-3′

Tip60 R: 5′-GGAAACACTTGGCCAGAAGA-3′

Ep400 F: 5′-CAGCTCCTCCTAAGCCACAG-3′

Ep400 R: 5′-CCTCTTGAAGCTTTGGCAAC-3′

FACS staining and sorting neuronal nuclei for DNA isolation

To sort neuronal nuclei for DNA isolation, dissected hippocampal tissue was placed in 1 ml of buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine and 0.5 mM spermidine) and dounced 5× with a loose pestle and 10× with a tight pestle. IGEPAL CA-630 (5%, 32 µl) was added before douncing again with a tight pestle 5–8 times. The nuclei suspension in HB was filtered through a 40-µm strainer. Nuclei were then pelleted by centrifugation at 500*g* for 5 min and resuspended in 400 µl of FACS block/stain buffer (1% BSA, 0.05% IGEPAL CA-630, 3 mM $MgCl_2$ in 1× PBS). Nuclei were incubated for 15 min with gentle rotation at 4 °C. Following this blocking step, nuclei were pelleted and resuspended in FACS block/stain buffer containing 1:1,000 dilution of mouse anti-NeuN-Alexa488 (Millipore, MAB377X, cloneA60). To set sort gates, an isotype control mouse IgG-Alexa488 (Life Technologies, MA518167) was included as a negative control along with an unstained sample. Samples were incubated in antibody mix for 1 h with gentle rotation at 4 °C. Nuclei were washed 1× with FACS block/stain buffer, and DRAQ5 nuclear dye (Abcam, ab108410) was added (1:500) before sorting. NeuN high-expressing nuclei were separated from NeuN low-expressing nuclei using a SONY SH800. Nuclei were sorted directly into ATL lysis buffer provided in a QIAamp DNA Micro kit. DNA was extracted using a QIAamp DNA Micro kit (Qiagen, 56304) according to the manufacturer's protocol and stored at −80 °C until amplicon library preparation. FACS analyses in gating figures shown throughout the manuscript were performed using FlowJo (10.0.8rl).

FACS sorting AAV-infected neuronal nuclei for CUT&RUN

Dissected hippocampal tissue was examined under a fluorescent scope to detect GFP (ΔCre) or mCherry (Cre). Tissue that was uninfected, or in rare cases showed infection of both fluorophores in a single hemisphere, was discarded. Hippocampi were placed in 0.5 ml of buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine and 0.5 mM spermidine) and dounced 5× with a loose pestle and 10× with a tight pestle. IGEPAL CA-630 (5%, 32 µl) was added before douncing with a tight pestle 5–8 more times and filtering through a 40-µm strainer. DRAQ5 nuclear dye (Abcam, ab108410) was added (1:500), and nuclei expressing either mCherry or GFP were sorted on a SONY SH800.

Negative gates were determined using uninfected tissue. Nuclei were collected in 1 ml of CUT&RUN wash buffer containing 2 mM EDTA. FACS analyses in gating figures shown throughout the manuscript were performed using FlowJo (10.0.8rl).

RNA-seq library preparation

Neurons infected with either control shRNA virus or viruses targeting *Npas4*, *EP400* or *Tip60* were washed twice with PBS to remove dead cells and scraped immediately into TRIzol. For wild-type time course in hippocampus paired with sBLISS-seq, microdissected tissue was flash frozen and then thawed in TRIzol. RNA was extracted using a RNAeasy kit (Qiagen) according to the manufacturer's instructions. Total RNA (1,000 ng) was used to generate libraries following ribosomal RNA depletion (NEBNext, E6310X) according to the manufacturer's instructions (NEBNext, E7420). For cultured neurons, 85 bp reads were generated on an Illumina NextSeq 500 and subsequently analysed with our standardized RNA-seq data analysis pipeline (below). For wild-type time course samples, 40 bp paired-end reads were obtained on an Illumina NextSeq 500.

ATAC-seq library preparation

For ATAC-seq libraries generated from hippocampal neuronal subtypes in vivo, *CamkIIa*-expressing CA1 pyramidal neurons were isolated from $Camk2a^{Cre};Sun1^{fl/+}$ mice using the INTACT method[53]. In brief, hippocampi were dounced 15× in buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine and 0.5 mM spermidine) to release nuclei. Nuclei were purified by spinning through an iodixanol gradient at 10,000*g* (see description for nuclear isolation in CUT&RUN experiments). Nuclei expressing SUN1–GFP on the nuclear membrane in a Cre-dependent manner were isolated by incubating the nuclear suspension with 10 µg of anti-GFP antibody (Invitrogen, G10362) for 30 min at 4 °C. Antibody-coated nuclei were subsequently captured by incubation with magnetic Protein G Dynabeads (Thermo Fisher). Following nuclear isolation and counting, approximately 30,000–40,000 nuclei per condition were resuspended in L1 lysis buffer (100 mM HEPES-NaOH pH 7.5, 280 mM NaCl, 2 mM EGTA, 2 mM EDTA, 0.5% Triton X-100, 1% NP-

40 and 20% glycerol), followed by 5 min incubation in L2 lysis buffer (10 mM Tris-HCl pH 8.0, and 200 mM NaCl) and 5 min incubation in ATAC lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630).

Nuclei were transposed using a Nextera DNA Library Prep kit (Illumina, FC-121-1030) as previously described[56]. Transposition was carried out for 30 min at 37 °C. Transposed DNA fragments from individual samples were purified, independently barcoded and amplified for 8–11 cycles. ATAC-seq libraries were selected for fragments ranging from 200 to 1,000 bp by gel electrophoresis and sequenced on an Illumina NextSeq 500 with 75 bp single-end reads.

CUT&RUN

CUT&RUN experiments were performed as previously described[6]. For mapping of NuA4 components in 4–6-week-old wild-type mice (Fig. 1), replicates consisted of pools of 4–5 mice processed separately. See Supplementary Table 2 for replicate numbers. Males and females were used in equal proportions in pooled replicate samples. In brief, fresh hippocampal tissue pooled from 4–5 mice was placed into 5 ml of buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine and 0.5 mM spermidine) and dounced 15× with a tight pestle. IGEPAL CA-630 (5%, 320 μl) was added before douncing with a tight pestle 5 more times, and the sample was filtered through a 40-μm strainer into a 15 ml conical collection tube. Five millilitres of working solution (50% iodixanol, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, supplemented with protease inhibitors,DTT, spermine and spermidine) was added to the sample. Homogenized tissue was gently added on top of 2 ml of 30/40% iodixanol layers. Samples were centrifuged at 10,000*g* for 18 min, and 1 ml of nuclei was collected from the 30/40% iodixanol interface. Nuclei were counted and evenly distributed across unstimulated and stimulated conditions in the following amounts: FOS (in house), H3K27ac (Abcam, ab4729): 75,000 nuclei; NPAS4 (in house), EP400 (Bethyl Laboratories, A300-541A; Abcam, Ab5201), ARNT2 (in house), RAD50 (Novus Biologicals, NB100-154): 1,000,000 nuclei; MRE11 (Novus Biologicals, NB100-142), CTCF (Active Motif, 61311), ETL4

(Bethyl Laboratories, A304-928A): 500,000 nuclei. For samples FACS-isolated from *Npas4* cKO mice, MRE11 and EP400 CUT&RUN was performed with 100,000 nuclei. Replicates of infected, FACS-sorted samples consisted of independent pools of two to three mice.

Equal numbers of nuclei between unstimulated (PBS-injected mice) and stimulated conditions (2 h KA-injected mice) were aliquoted into 1 ml of CUT&RUN wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.2% Tween-20, 1 mg ml^{-1} BSA, 10 mM sodium butyrate and 0.5 mM spermidine supplemented with protease inhibitors). Magnetic concanavalin-A (ConA) beads (Bangs Laboratories) that had been washed with CUT&RUN binding buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$ and 1 mM $MnCl_2$) were added to each sample to bind nuclei. ConA-bead-bound nuclei were incubated overnight at 4 °C in wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.2% Tween-20, 1 mg ml^{-1} BSA, 0.1% Triton X-100, 2 mM EDTA, 10 mM sodium butyrate and 0.5 mM spermidine supplemented with protease inhibitors) containing 1:50 dilution of the aforementioned antibodies.

After overnight incubation with antibodies, ConA-bead-bound nuclei were washed with CUT&RUN antibody buffer and resuspended in CUT&RUN Triton-wash buffer (CUT&RUN wash buffer supplemented with 0.1% Triton X-100). Protein-A-MNase was added at a final concentration of 700 ng ml^{-1}, and samples were incubated at 4 °C for 1 h. ConA-bead-bound nuclei were next washed twice with CUT&RUN Triton-wash buffer and resuspended in 100 µl of CUT&RUN Triton-wash buffer. After the addition of 3 µl of 100 mM $CaCl_2$, samples were incubated on ice for 30 min. To stop the MNase reaction, 100 µl of 2× STOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.04% Triton X-100, 20 pg ml^{-1} yeast spike-in DNA and 0.1 µg ml^{-1} RNase A) was added to each sample before incubation at 37 °C for 20 min. Following incubation, a magnet was used to capture Con-A-beads, and supernatants containing DNA fragments released by protein-A-MNase were collected. Two microlitres of 10% SDS and 2 µl of 20 mg m^{l-1} proteinase K were added to supernatants followed by incubation at 65 °C with gentle shaking for 1 h. Standard phenol–chloroform extraction with ethanol precipitation was used to precipitate DNA.

Sequencing libraries from precipitated DNA suspend in TE buffer were generated as previously described[57], with the following changes: Rapid T4 DNA ligase (Enzymatics) was used to perform adapter ligation onto end-repaired and A-tailed DNA. Adaptor dimers were removed from PCR-amplified libraries using a 1.1× ratio of AMPure XP beads. CUT&RUN libraries were sequenced on Illumina NextSeq 500 using 40-bp paired-end reads.

ChIP–seq

For ChIP–seq, replicates consisted of pools of four to five mice performed on independent days. For ChIP–seq of NPAS4 from hippocampal tissue, hippocampi from 15 mice were dounced in 5 ml of 1× PBS containing protease inhibitor cocktail (Roche, 11836153001). Formaldehyde (1%) was added to tissue homogenate and incubated for 10 min at room temperature, followed by the addition of 0.125 M glycine for 5 min at room temperature. For ChIP of γH2AX from *Camk2a*-expressing CA1 pyramidal neurons, nuclei were isolated from $Camk2a^{cre};Sun1^{fl/+}$ mice using the INTACT method[53] (see ATAC-seq library preparation section for details of nuclear isolation). Following isolation, nuclei with attached beads were crosslinked with 1% formaldehyde in 1 ml of 1× PBS for 10 min at room temperature. Crosslinking was quenched by the addition of 0.125 M glycine for 5 min at room temperature. Crosslinked nuclei were frozen and stored at −80 °C before proceeding with the protocol outlined below.

Nuclei were release from tissue by 10 min incubation in lysis buffer 1 (LB1) (100 mM HEPES-NaOH pH 7.5, 280 mM NaCl, 2 mM EDTA, 2 mM EGTA, 0.5% Triton X-100, 1% NP-40 and 20% glycerol) followed by washing in buffer containing 10 mM Tris-HCl pH 8.0, and 200 mM NaCl. Chromatin was sheared using a Bioruptor (Diagenode) on high power mode for 40–42 cycles with 30-s pulses in sonication buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate and 0.5% *N*-lauroylsarcosine).

Following sonication, 1.5 ml of chromatin from hippocampal tissue (about 60 µg) was supplemented with 1% Triton and incubated overnight at 4 °C with 4 µg NPAS4 (in house) coupled to 15 µl Protein A Dynabeads

(Invitrogen, 10001D). For γH2AX ChIP, 1.5 ml of chromatin released from around 100,000 purified nuclei was incubated with 2 µl of anti- γH2AX (Abcam, ab2893) coupled to 15 µl Protein A Dynabeads. Beads were washed twice sequentially in 0.5 ml of the following buffers: low-salt buffer (20 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), high-salt buffer (20 mM Tris pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), LiCl wash buffer (10 mM Tris pH 8, 1 mM EDTA, 1% NP-40, 250 mM LiCl and 1% sodium deoxycholate), followed by a wash in 1 ml of TE buffer. Protein–DNA complexes were eluted off the beads by incubation with 200 µl of TE plus 1% SDS for 30 min at 65 °C. Crosslinked DNA was reversed by incubation overnight at 65 °C. The following day, RNA and protein were digested away by the addition of 10 µg of RNase A and 5–7 µl of proteinase K (New England Biolabs, P8107S). DNA was purified by standard phenol–chloroform extraction. ChIP–seq libraries were generated using an Ovation Ultralow V2 kit (Nugen, 0344-32) according to the manufacturer’s instructions and PCR amplified for 13–16 cycles, depending on the antibody. Library quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). Reads (75 bp) were generated using an Illumina NextSeq 500.

snRNA-seq

To isolate hippocampal nuclei, we placed flash-frozen hippocampal tissue in 0.5 ml of buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine and 0.5 mM spermidine) and dounced 5× with a loose pestle and 10× with a tight pestle. IGEPAL CA-630 (5%, 32 µl) was added before douncing with a tight pestle 5–8 more times and filtering through a 40-µm strainer into a 15 ml conical collection tube. Buffer HB (3.5 ml) and 5 ml working solution (50% iodixanol, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, supplemented with protease inhibitors, DTT, spermine and spermidine) were added. Homogenized tissue was gently layered on top of 1 ml of 30% iodixanol on top of a layer of 1 ml of 40% iodixanol (diluted from a working solution). Samples were centrifuged at 10,000*g* for 18 min, and 70 µl of nuclei was collected from the 30/40% iodixanol interface. An aliquot of each sample was incubated with trypan blue, and nuclei were counted using a standard haemocytometer.

snRNA-seq was performed using a 10x Genomics Chromium Single Cell kit (v.3). Each reaction lane was loaded with up to 10,000 nuclei from one hippocampal hemisphere (infected with either Cre-mCherry or ΔCre-GFP) from one mouse. Subsequent steps for cDNA amplification and library preparation were conducted according to the manufacturer's protocol (10x Genomics). Samples were sequenced using an Illumina NextSeq 500 with 28 bp (R1), 56 bp (R2) and 8 bp (index) reads.

To facilitate detection of viral transcripts, 2 µl of remaining UMI-barcoded cDNA was amplified in a separate set of PCRs to increase the abundance of UMI-labelled viral transcripts. Custom primers were used to amplify mCherry or GFP transcripts present in the barcoded cDNA library. A nested PCR strategy using Q5 high-fidelity polymerase was used to reduce nonspecific products, with the forward primer in the second reaction also adding the sequence for Illumina Read2 (R2) to the amplified product (primers: GFP amplification 1: 5′-CGCCGACCACTACCAGCAGAACACC-3′; GFP amplification 2: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgctgganttcgtgaccgccgcc-3′; mCherry amplification 1: 5′-CACTACGACGCTGAGGTCAAGACCACC-3′; mCherry amplification 2: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgcgccgagggccgccactcc-3′). PCR products were isolated following the first and second steps of the nested PCR using a 0.6× size selection with SPRIselect reagent. Final products from the nested PCR were then diluted 1:10 and put into a sample indexing PCR using the sample indexing primer and Chromium i7 sample index plate, as described in the 10x Genomics library preparation protocol. These sample-indexed viral transcripts were spiked into the sequencing runs used for the corresponding full cDNA libraries. Our final datasets include 32,418 nuclei collected from two independent *Npas4*$^{fl/fl}$ mice and 44,511 nuclei collected from three independent *Tip60*$^{fl/fl}$ mice. Ultimately, we identified 12,963 Cre-infected and 8,845 ΔCre-infected nuclei in the *Npas4*$^{fl/fl}$ snRNA-seq dataset, as well as 13,536 Cre-infected and 13,461 ΔCre-infected nuclei in the *Tip60*$^{fl/fl}$ dataset (Extended Data Fig. 4).

sBLISS-seq in cultured neurons and in isolated hippocampal nuclei

Nuclear isolation, fixation and adapter ligation

sBLISS-seq was carried out on cultured neuronal nuclei or on nuclei isolated from hippocampal tissue according to a previously described protocol[30]. Replicates consisted of individual mice for the wild-type time course. For the Cre and ΔCre datasets, a replicate consisted of one infected hemisphere from one mouse. Samples were paired such that 2 h KA_CRE1 and 2 h KA_ΔCre1 came from the same mouse. For sBLISS-seq performed on *Npas4* or *Tip60* wild-type or cKO nuclei infected with AAVs, hippocampal tissue was dissected and quickly examined under a fluorescent scope to detect GFP (ΔCre) or mCherry (Cre). Tissue that was uninfected, or in rare cases showed infection of both fluorophores in a single hemisphere, was discarded. For AAV-infected hippocampi, the tissue was not sorted to minimize stress on the nuclei that could induce ectopic DNA breaks. For uninfected hippocampal nuclei used in the wild-type sBLISS-seq time course, the hippocampi were dissected and processed as described below.

To process hippocampal tissue for sBLISS-seq, hippocampi were placed in 1 ml of buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine and 0.5 mM spermidine) and dounced 5× with a loose pestle and 10× with a tight pestle. IGEPAL CA-630 (5%, 32 µl) was added before douncing with a tight pestle 5–8 more times and filtering through a 40-µm strainer. To preserve DNA breaks for the remainder of the protocol, nuclei were fixed with 2% PFA (Electron Microscopy Sciences, 15710) followed by quenching with 125 mM glycine. Fixed nuclei were gently pelleted by centrifuging at 500*g*. To remove excess debris and myelin, fixed nuclei were run through a modified iodixanol gradient. Pelleted nuclei were resuspended in 1 ml of 22% iodixanol, laid on top of 100 µl of 43% iodixanol. Nuclei were then centrifuged at 1,500*g* for 15 min in an Eppendorf Centrifuge 5804R in a swinging bucket rotor. Nuclei (100 µl), which settle at the 43/22% interface, were collected and transferred to a Protein Low-Bind tube. Nuclei were washed once with ice-cold PBS plus 1% BSA followed by 2 washes with ice-cold 1× PBS. Fixed nuclei were counted (715,000 nuclei for the wild-type time course, around 250,000 nuclei for *Npas4*$^{fl/fl}$ and wild-type Cre and ΔCre datasets, and about 100,000 nuclei for *Tip60*$^{fl/fl}$ Cre and ΔCre datasets) and were transferred to a fresh tube to begin adapter ligation.

Note that the wild-type time course samples were performed in two batches with stimulation time points split across these two batches. We did observe a batch effect from processing the samples, which was computationally removed (see analysis of sBLISS-seq samples). The batch for the associated samples is listed in Supplementary Table 2. Note also that the wild-type time course samples included a 2% spike-in of human HEK293T cells expressing a guide-RNA-induced cut site. Spike-in normalization was not performed in the final data processing owing to low expression of the guide and low coverage of reads across the cut site per sample.

To isolate nuclei from cultured neurons plated in 6-well dishes for Cas9 control experiments, neurons were washed with ice-cold PBS to remove debris and 1 ml of HB, and 32 µl of 5% IGEPAL CA-630 was added to each well. Cells were incubated for 10 min in HB with gentle rotation at 4 °C before removal by gentle scraping and transfer to Eppendorf tubes. Nuclei from cultured neurons were fixed with 2% PFA (Electron Microscopy Sciences, 15710) followed by quenching with 125 mM glycine and gently pelleted by centrifuging at 500*g*. The aforementioned gradient was omitted owing to a lack of debris and myelin in neuronal cultures. Nuclei were washed once with ice-cold PBS plus 1% BSA followed by 2 washes with ice-cold 1× PBS. Roughly 1×10^6 cultured nuclei were used for in vitro CRISPR–Cas9 datasets.

Fixed nuclei were incubated for 1 h at 37 °C in LB2 (10 mM Tris-HCl, 150 mM NaCl and 0.3% SDS). Nuclei were washed twice in CutSmart Wash Buffer (1× CutSmart (New England Biolabs, B7204) and 0.1% Triton X-100). At room temperature, nuclei were blunted using a Quick Blunting kit (New England Biolabs, E1201) for 1 h. Following two washes with CutSmart wash buffer, adapters were ligated onto free DNA ends in intact nuclei using T4 DNA ligase (5 U ml^{-1}; Thermo Fisher Scientific, EL0011). Ligation reactions were carried out for about 18 h at 16 °C in a Thermomixer. The following day, nuclei were washed twice with CutSmart wash buffer and incubated overnight in 100 µl of DNA extraction buffer (10 mM Tris-HCl, 100 mM NaCl, 10 mM EDTA and 0.5% SDS) containing 10 µl of proteinase K (20 mg ml^{-1}; New England Biolabs). DNA was extracted using standard phenol–chloroform extraction and resuspended in 18 µl of ultrapure DNase/RNase-free water for subsequent fragmentation.

DNA fragmentation, IVT and library production

In vitro and wild-type time course samples were fragmented using a Bioruptor, 30 s on, 60 s off, high intensity, for 35–40 cycles. Owing to limited input from the Cre and ΔCre samples, we found that using enzymatic fragmentation preserved fragmented DNA better than sonication. As a result, all samples from hippocampal tissue were fragmented using NEB Fragmentase for 40–50 min at 37 °C. This treatment resulted in fragments 400–800 bp in size, with the average library size approximately 650 bp. Following fragmentation and clean-up, 100 ng (for cultured neuron samples), 200 ng (for wild-type time course samples) or 35 ng (Cre and ΔCre samples) was reverse transcribed using a MegaClear IVT kit (Thermo Fisher, AMB13345). Libraries from IVT products were produced as previously described without deviation from the protocol[30]. In brief, template DNA was removed by incubating the IVT products with 2 μl of DNaseI (Thermo Fisher, AM2222) for 15 min, and RNA was purified using RNAXP clean beads. RA3 adapter (/5rApp/TGGAATTCTCGGGTGCCAAGG/3SpC3/) was ligated onto RNA using T4 RNA Ligase 2, truncated (New England Biolabs, M0242L). RNA was then reverse transcribed using the reverse primers RTP (5′-GCCTTGGCACCCGAGAATTCCA-3′) and a SuperScript IV Reverse Transcriptase kit (Thermo Fisher,18090200). Finally, DNA was amplified using NEBNext Ultra II Q5 master mix (New England Biolabs, M0544L) in 8 reactions of 50 μl each. Libraries were sequenced using single-end 75 bp reads on an Illumina NextSeq 500.

END-seq

Culture mouse cortical neurons for END-seq were infected with a non-targeting shRNA lentivirus on 3 DIV before collection on 7 DIV. Replicates consisted of cultures generated on independent days. To dissociate cultured mouse cortical neurons for END-seq, papain (Worthington Biochemical, LK003178) was dissolved in TrypLE Express enzyme solution at 37 °C for 20 min before sample collection. Culture medium was gently aspirated and the cells were gently washed once with PBS at 37 °C. Papain solution (500 μl) was added to each well of neurons in 6-well plates and incubated at 37 °C for 1 min. Papain solution was removed using a pipette, and 500 μl

trituration solution (culture medium supplemented with freshly dissolved DNase) was added. Cells were gently triturated 5–10 times with a wide bore pipette tip, transferred to conical tubes and pelleted. Cells were gently resuspended in ice-cold PBS containing 0.1% BSA and 0.5 mM EDTA for counting. Cells were stored on ice while an aliquot was quickly counted using a haemocytometer. Next, 1.5 million neurons for each experimental condition were added to a new conical tube and pelleted. Cells were then embedded in agarose plugs using a CHEF Mammalian Genomic DNA Plug kit (Bio-Rad, 1703591) and processed for END-seq as previously described[32,58]. Experimental groups were processed back-to-back, collecting cells in one treatment group and embedding them in agarose plugs before collecting cells in the next treatment group to reduce overdigestion of the samples and to minimize time between cell dissociation and embedding in agarose. We observed better signal-to-noise in cells treated with etoposide (50 μM). Etoposide was added to the cells 6 h before collection and was included at 50 μM in all solutions applied to the cells until embedding in agarose.

Amplicon PCR sequencing (SiMSen-seq)

Selection of sites for mutation analysis

To select sites to examine for mutational load during ageing, a universe of possible sites to examine was created by taking the union of all accessible regions in *CamkIIa*-expressing neurons (the predominant cell class in CA1 hippocampal regions; defined in our INTACT *CamkIIa* datasets) and all regions marked by H3K27ac in hippocampal nuclei. This generated a set of 179,841 possible regions, and we considered windows of 500 bp from the centre of these peaks. We calculated the normalized read intensity for ATAC-seq, H3K27ac CUT&RUN and γH2AX ChIP–seq using the function in DeSeq2 counts(dds, normalized=TRUE)[59] for each of these possible regions across all conditions. We also extracted the normalized signal intensity for NPAS4 and MRE11 CUT&RUN using Homer(homer/4.9) annotatePeaks.pl function with default normalization to 10 million total sequencing reads. Finally, we determined regions that overlapped a NPAS4 and/or MRE11 peak by at least 1 bp. These sites were considered ‘bound’.

Using the previously mentioned set of regions, we chose sites to interrogate based on the following criteria: (1) sites bound by NPAS4 and MRE11 with the highest normalized signal for NPAS4; (2) sites bound by NPAS4 and MRE11 that show an increase in γH2AX ChIP–seq signals following neuronal activation. We also included a set of sites that did not overlap with a NPAS4 or MRE11 peak and as a group did not differ significantly in terms of their GC content, chromatin accessibility or levels of H3K27ac. However, we note that in our final site selection, some sites that matched these criteria did not amplify efficiently in our assay and were therefore excluded owing to technical considerations. The final set of sites assayed are found in Supplementary Table 5.

Primers were designed using custom R scripts calling Primer3, with Primer3 formatting derived from http://bioinfo.ut.ee/primer3-0.4.0/input-help.html. The following sequence was appended onto the 5′ end of the forward primer GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNATGGGAAAGAGTGTCC, where the 12 Ns represent random nucleotides constituting a UMI. The following sequence was appended onto the 5′ end of the reverse primer GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT. Primer sequences for the amplicons targeted can be found in Supplementary Table 5. Primers were ordered as 200 pM or 4 nM Ultramers from IDT.

Production of targeted amplicon libraries

Amplicon libraries were generated using a previously described protocol[38] with minor modifications. For all samples, sites of interest were divided into pools of 15 primer sets (see Supplementary Table 5 for primer pooling). For each primer pool, 20–30 ng of DNA isolated from the FACS-isolated $NeuN^+$ nuclei were first amplified with Phusion Hot Start II high-fidelity DNA polymerase (Thermo Fisher, F-549L) to append UMI sequences to DNA templates. The final concentration of each primer in the reaction was 40 nM. PCR parameters for these 4 cycles were as follows: 98 °C, 30 s; 4 cycles of 98 °C, 10 s, 62 °C, 6 min 72 °C, 30 s; followed by 65 °C 15 min; 95 °C, 15 min. During the 65 °C incubation, the PCR reaction was terminated by incubation with 10 µl of protease (Sigma-Aldrich, P5147-100MG; used at a final concentration of 0.06 µg μl^{-1}). For each primer pool, 10 µl of initial

product were next amplified using Phusion Hot Start II high-fidelity DNA polymerase with the following parameters: 98 °C, 3 min; 2–35 cycles of 98 °C, 10 s 80 °C, 1 s; 72 °C, 30 s; 76 °C, 30 s. All with ramping at 0.2 °C s^{-1}. Duplicate reactions were performed for each primer pool to increase the diversity of UMI sampling in the final libraries. Samples were size-selected to include products approximately 275–650 bp in length. Analysis was conducted using Debarcer v.0.3.1 (https://github.com/oicr-gsi/debarcer/releases/tag/v0.3.1)[38].

Lifespan analysis and tissue collection of NPAS4 knockout cohort

Npas4 wild-type or knockout littermates were housed in accordance with protocols approved by the Harvard University Standing Committee on Animal Care. Littermates were housed together and monitored by trained technicians for overall health. Any distressed animals or those that showed poor health were euthanized and censored during data collection. Animals that were used for tissue collection purposes before the final collection date were censored from the final lifespan analysis. Owing to limitations imposed by the COVID-19 shutdown, the final study ended in March 2020, when all remaining animals were dissected. For this reason, the lifespans of *Npas4* wild-type littermates do not go to completion. Significance of survival curves was determined using both a log-rank (Mantel–Cox) and a Gehan–Wilcoxon–Breslow test using Prism (v.8.4.2) software.

Quantification and statistical analysis

Statistical analysis and sample size determination

The statistical analysis for each experiment is detailed in the figure legends. Electrophysiological assays were performed in a blinded manner such that the condition (saline-injected versus KA-injected) was not revealed until after the analysis was complete. Statistical methods were not used to predetermine sample sizes, but replicate numbers generally adhered to guidelines of the ENCODE consortium[60]. Sample randomization was not performed. All statistical analysis was performed in either Prism (v.8.4.2) or R (v.3.6.1). For multiple hypothesis correction shown in Extended Data

Figs. 6e and 13b, *P* values obtained using Prism (v.8.4.2) were corrected using the padjust() function in base R (v.3.6.1).

Peptide quantification mass spectrometry

Mass spectrometry and peptide quantifications were performed following the standard practices of the Taplin Biological Mass Spectrometry facility, Harvard Medical School. Data were collected using a LTQ Orbitrap Velos Elite ion-trap mass spectrometer (Thermo Fisher Scientific). In brief, the program Sequest (Thermo Fisher Scientific) was used to compare peptides against protein databases with the acquired fragmentation pattern. Data were filtered to between a 1% and 2% peptide false discovery rate (FDR), and databases included a reversed version of all the sequences. Triplicate NPAS4 immunoprecipitation–mass spectrometry experiments were performed from hippocampal tissue. Duplicate experiments were performed on high molecular weight fractions using NPAS4–Flag-HA and TIP30–H3-Flag cortical lysates. Peptide counts for all mass spectrometry experiments are provided in Supplementary Table 1. The R (3.6.1) package EdgeR[61] (edgeR_3.28.1;limma_3.42.2) was used to identify proteins significantly enriched in NPAS4 or TIP60 immunoprecipitate samples relative to wild-type samples that did not express Flag-tagged proteins. Proteins that were identified in at least 2 out of 3 replicates were included in the EdgeR analysis. Peptides found strictly in wild-type control samples and not in NPAS4 or TIP60 samples (that is, background associated with the M2 resin) were removed before running the EdgeR glmFit() and glmLRT() functions.

Sequencing and alignment

All experiments were sequenced on an Illumina NextSeq 500 (Illumina Next-Seq Control Software v.4.0.2). Information on sequencing data is provided in Supplementary Table 2. Single-end reads (75 bp) were obtained for ATAC-seq, ChIP–seq, RNA-seq (cultured neurons) and sBLISS-seq. Paired-end reads (40 bp) were obtained for all CUT&RUN experiments, snRNA-seq experiments and hippocampal KA time course RNA-seq experiments. Single-end reads (162 bp) were obtained for amplicon libraries used in the mutation analysis.

For ATAC-seq and ChIP–seq samples (that is, NPAS4 ChIP and γH2AX ChIP), quality trimming of sequencing reads was performed with trimmomatic/0.36 (ref. [62]) using the following command: java -jar trimmomatic-0.33.jar SE -threads 1 -phred33 [FASTQ_FILE] ILLUMINACLIP:[ADAPTER_FILE]:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:35. Nextera adapters were specified for ATAC-seq, and TruSeq adapters were specified for ChIP–seq samples. Samples were subsequently aligned to the mm10 genome using the Bowtie alignment software (vbowtie2/2.2.9) with the –very-sensitive setting. Reads mapping to the mitochondrial genome were removed. Duplicate reads were marked with Picard/2.8.0 with the command java -jar $PICARD/picard-2.8.0.jar MarkDuplicates REMOVE_DUPLICATES=false. Duplicates were subsequently removed using samtools/1.3.1 samtools view -b -F 1796.

For CUT&RUN samples, quality read trimming was performed using trimmomatic/0.36 using the following command: java -jar trimmomatic-0.33.jar PE -phred33 [FASTQ_FILE] ILLUMINACLIP:/n/app/trimmomatic/0.36/bin/adapters/TruSeq3-PE.fa:2:15:4:4:true LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:25. Adapters were further trimmed according to the pipeline and script (kseq_test) as previously described[63]. Paired-end reads were mapped to the mouse genome using bowtie2/2.2.9 with the following command: bowtie2 --local --very-sensitive-local --no-unal -x mm10 --dovetail --no-mixed --no-discordant --phred33 -I 10 -X 700. Yeast spike-in reads were mapped using the following command: bowtie2 --local --very-sensitive-local --no-unal –x sacCer3 --no-overlap --no-dovetail --no-mixed --no-discordant --phred33 -I 10 -X 700. Read depth (normalized to 1 million) or spike (normalized to 1) normalized bedgraph files were created using custom scripts modified from Spike_in_Calibration_v2.csh as previously described[63]. Finally, bedGraphToBigWig (ucsc-tools/363) was used to generate the bigWig files from either read depth normalized bedgraph files. Files are displayed on the IGV browser show read depth-normalized bigWigs. Mapping for sBLISS-seq samples were performed according to a previously described protocol[30]. The mapping pipeline was cloned from git clone at https://github.com/BiCroLab/blissNP.git.

For amplicon library analysis, adapter contamination was a significant problem for $NeuN^{+}$-sorted datasets. Samples were therefore trimmed to remove these adapters with the following trimming command: java -jar trimmomatic-0.33.jar SE -threads 1 -phred33 [FASTQ_FILE] ILLUMINACLIP:[ADAPTER_FILE]:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:120. FASTQ files containing only trimmed, adapter-removed reads are provided in the Gene Expression Omnibus (GEO) submission. Full FASTQ files will be provided upon request. Mapping using BWA was performed using the default parameters outlined in the package Debarcer v.0.3.1 (https://github.com/oicr-gsi/debarcer/releases/tag/v0.3.1)[38].

END-seq reads were aligned to reference genome mm10 using subread (v.1.5.1) with parameters subread-align -t 1 -T 6 -M 3. The Samtools (v.1.13) functions sort, markdup and index were used to create indexed .bam files with duplicate reads removed. Files were downsampled to the lowest number of reads in any sample within a replicate, and tag directories were compiled using the homer (v.4.9) command makeTagDirectory. Tag directories were used to generate aggregate plots centred on various genomic sites of interest using homer annotatePeaks.pl with parameters -size 10000 -hist 50. For statistical tests, signals were extracted within windows around the genomic sites of interest using homer annotatePeaks.pl with parameters -size 500.

SAR-seq analysis

Previously published[32] SAR-seq datasets were downloaded from the GEO database, accession number GSE167259. Three replicates of SAR-seq performed in iNeurons (postmitotic glutamatergic neurons derived from induced pluripotent stem cells) were retrieved (GSM5100400, GSM5100401 and GSM5100402). Tag directories were used to generate aggregate plots centred on various genomic sites of interest using homer annotatePeaks.pl with parameters -size 10000 -hist 50. For statistical tests, signals were extracted within windows around the genomic sites of interest using homer ‘annotatePeaks.pl’ with parameters -size 500. Site lists generated in mice (mm10) were lifted over to human hg19 for use with human SAR-seq data, using the UCSC Genome Browser liftOver tool. To generate sites with

nonoverlapping signals for NPAS4 and FOS, NPAS4 and FOS summits were both extended to 1 kb. NPAS4-bound sites (no FOS) were generated using bedtools/2.27.1 intersect bed -v option to generate peaks with no NPAS4 and vice versa.

bigWig visualization

To generate bigWig files for ATAC-seq, ChIP–seq and CUT&RUN datasets, all aligned bam files for each replicate of a given experimental condition were pooled and converted to the BED format with the bedtools/2.27.1 bamtobed. For ATAC-seq and ChIP–seq data, the 75 bp reads were extended in the 3′ direction to 200 bp (average fragment length for ChIP–seq and ATAC-seq experiments as measured by bioAnalyzer) with the bedtools slop command using the following parameters: -l 0 -r 125 -s. For sBLISS-seq libraries, the mapping pipeline generates a single base pair cut site. Thus, for visualization purposes, reads were extended 75 bp in the 3′ direction to match the initial read sequence. Mm10 blacklisted regions were filtered out using the following command: bedops –not-element-of 1 [BLACKLIST_BED][60]. The filtered BED files were converted to coverageBED format using the bedtools genomecov command with the following options: -scale [NORM_FACTOR to scale each library to 20M reads for ChIP, 20M for ATAC, 10M for sBLISS-seq, and 1M for CUT&RUN] –bg. Finally, bedGraphToBigWig (ucsc-tools/363) was used to generate the bigWig files displayed on browser tracks throughout the manuscript using the IGV browser.

Bulk RNA-seq quantification of gene expression in vitro RNA-seq

The featureCounts package[64] was used to count reads in cultured neuron RNA-seq data using a custom filtered annotation file (gencode.v17.annotation.gtf filtered for feature_type="gene", gene_type="protein_coding" and gene_status="KNOWN") to obtain read counts along genes for each sample. For differential gene expression analysis, read count tables were TMM-normalized using the EdgeR software analysis package[61]. Any genes that were not expressed in at least three samples with TMM-normalized CPM > 1 were removed from further analysis. The voom and limma (edgeR_3.28.1;limma_3.42.2) analysis

software packages were used to quantify differential gene expression (requiring FDR-corrected $q < 0.01$). For analysis of the paired RNA-seq samples that matched sBLISS-seq, the DeSeq2 package was used to generate normalized counts. After running a standard DeSeq2 pipeline, normalized counts were generated with the function counts(dds, normalized=TRUE)[59].

Gene ontology enrichment analysis

Gene ontology (GO) enrichment analysis was performed using gProfiler2 in R (v.0.2.0)[65], with a custom background of expression-filtered genes from neuronal cell types across all scRNA-seq datasets in this manuscript and FDR < 0.05. Select GO terms are displayed in Extended Data Fig. 5e and a complete list of enriched GO terms for each cell type and dataset can be found in Supplementary Table 3.

snRNA-seq analysis

FASTQ files were created using the standard bcl2fastq pipeline from Illumina. Gene expression tables for each nuclear barcode were generated using the CellRanger 3.0.0 pipeline as designed by 10x Genomics. Samples were demultiplexed, and all *Npas4*$^{fl/fl}$ or *Tip60*$^{\mathrm{fl/fl}}$ samples were merged using the CellRanger aggr function using default parameters. The datasets were loaded into R and analysed using the Seurat (v.3)[66] and Monocle3[67] packages. Nuclei were removed from the dataset if they contained fewer than 500 detected genes, displayed more than 5% of reads mapping to mitochondrial genes or had RNA counts detected at a level greater than 2 standard deviations higher than the average value in their assigned cell type (which probably reflect doublets and multiplets). To remove potential doublets from the datasets in a more stringent manner, we used the DoubletFinder package in R, which assesses which barcodes in a dataset are most likely to be doublets based on transcriptional similarity to distinct clusters[68]. We used default parameters and an estimated doublet rate of 7% based on guidance from 10x Genomics Chromium Single Cell 3′ Reagent kits v3 User Guide and the number of nuclei run in each reaction lane.

All nuclei in either *Npas4*$^{fl/fl}$ or *Tip60*$^{fl/fl}$ samples were considered together for clustering and dimensionality reduction. The 2,000 top variable genes across nuclei were identified using Seurat's (v.3) FindVariableFeatures function, which were used to perform PCA using the RunPCA function (both using default parameters). A shared nearest neighbour graph was constructed using the FindNeighbors function (considering the top 30 principal components), and clustering was assigned using the FindClusters function (resolution $n = 0.02$). The following marker genes were used to assign cell type to the principal ten clusters identified: pan-neuronal (*Rbfox3*); pan-excitatory neurons (*Slc17a7*); pan-inhibitory neurons (*Gad2*); excitatory dentate gyrus neurons (*Prox1* and *C1ql2*); excitatory CA3 neurons (*Cpne4* and *Spock1*); excitatory CA1 neurons (*Mpped1*); excitatory subiculum neurons (*Tshz2*); oligodendrocytes (*Mog* and *Mag*); oligodendrocyte precursor cells (*Pdgfra*); astrocytes (*Gfap*); microglia (*Cx3cr1* and *C1qc*); and endothelial cells (*Cldn5*) (Extended Data Fig. 4). To assign infection status to each nucleus, we set a threshold of detecting more than eight mCherry or GFP transcripts in a given nucleus, which represented an inflection point above the background rate of detection for the distribution of these transcripts per nucleus, and reflected the expected infection patterns based on the known tropism of the AAV2/9 virus used in these experiments. Differential gene expression analysis between Cre and ΔCre-infected cells within each cluster was conducted using Wilcoxon rank-sum test using the FindMarkers function (logfc.threshold = 0.25), and significant genes were defined as those with an adjusted $P < 0.01$. Heatmaps were generated using custom functions written in R. Violin plots were generated using Seurat VlnPlot function with default parameters.

We observed that in both datasets, within each neuronal cluster, the cells subclustered according to infection status (Cre versus ΔCre), and this effect was not driven by expression of viral transcripts, as all viral features were removed from the gene expression matrix. This subclustering probably reflects the inability of Cre-infected cells to fully induce activity-dependent genes following depletion of either *Npas4* or *Tip60*. At 6 h after stimulation, these activity-induced genes were among the most highly expressed genes and differed across neuronal cell types, and therefore contributed significantly to the principal components used in dimensionality reduction for cell-type clustering. We used two independent analysis packages, Seurat

(v.3) and Monocle3, to call differentially expressed genes and only considered significant genes (adjusted $P < 0.01$) from both analyses. Although we identified genes that are both upregulated and downregulated after acute *Npas4* or *Tip60* depletion, we focused our downstream analysis on downregulated genes that are more likely to be directly activated by this complex. To ensure our observed downregulation of putative target genes was not simply due to higher expression or detection of genes in the ΔCre-infected tissue, we randomly sampled from the top 10% of most highly expressed genes (to account for the target genes being highly expressed overall) and calculated the ΔCre – Cre difference for the set of randomly selected genes. We performed 10,000 iterations of this analysis to generate a distribution of sample differences using randomly selected genes. The actual observed differences (using the NPAS4–NuA4 target genes identified in each neuronal subtype) lay far outside this distribution (Extended Data Fig. 5c).

In general, NPAS4 induces a diverse set of pan-neuronal and cell-type-specific effector genes. Many targets were found to be common among 2 or 3 cell types, and about 10% of targets were specific to one cell class. Broadly, the set of all NPAS4 target genes were enriched for genes with functions in cell–cell adhesion, intercellular signalling and axon guidance, as well as a diverse set of metabotropic and ionotropic neurotransmitter receptor subunits. In addition to capturing known NPAS4 target genes such as *Nptx2*, *Plk2* and *Bdnf*, we identified 1,766 potentially new targets of NPAS4 in the hippocampus by this analysis (Supplementary Table 3).

ATAC-seq and ChIP–seq peak calling

ATAC-seq enriched peaks were determined using MACS2 (v.2.1.1) parameters –shift 100 -p 1e-5 --nolambda --keep-dup all --slocal 10000, as previously described[56,69]. ATAC-seq peaks from individual blacklist regions were removed as previously described[60]. sBLISS-seq peaks were called on UMI de-duplicated bed files in which the single cut site had been extended by the length of the sequencing read (75 bp). Peaks were called using MACS2 (v.2.1.1) with the following peak parameters: macs2 callpeak --nomodel --keep-dup all --format BED -g mm -p 1e-5. Reproducible peaks were considered those that overlap in 5 out of 8 replicates. Peaks mapping to the mitochondrial genome were removed. Final peak sets were extended 500 bp from the maximal sBLISS-seq summit in a reproducible peak. Peak calling on ChIP samples was performed using MACS2 (macs2/2.1.1)[70] using the following command macs2 callpeak –t (experimental bam) –c (input bam) -f BAM -g mm -p 1e-5.

CUT&RUN peak calling

All peak calling was performed using SEACR_1.1.sh[71]. For H3K27ac, FOS, CTCF and RAD50 CUT&RUN datasets, peak calling on individual replicates was performed using the spike-in normalized bedgraph files based on fragments 1–1,000 bp in length with the following command: SEACR_1.1.sh [target bedgraph] [control bedgraph] norm stringent. Paired control samples (either IgG or knockout control) are listed in Supplementary Table 2. To identify reproducible peak sets, SEACR peaks found in 3 out of 3 H3K27ac replicates (0 and 2 h KA stimulation), 3 out of 3 FOS replicates (0 and 2 h KA stimulation), 3 out of 3 CTCF replicates (0 and 2 h KA stimulation), 3 out of 3 RAD50 replicates (0 h) and 3 out of 4 RAD50 replicates (2 h KA stimulation) were intersected using bedtools/2.27.1 intersect bed. Peaks within 150 bp were merged. For NPAS4, ARNT2, EP400, MRE11 and ETL4 CUT&RUN datasets, peak calling on individual replicates was performed using SEACR_1.1.sh (ref. [71]) using the spike-in normalized bedgraph files based on fragments 1–1,000 bp in length with the following command: SEACR_1.1.sh [target

bedgraph] [control bedgraph] norm relaxed. Paired control samples (either IgG or knockout control) are listed in Supplementary Table 2. To identify reproducible peak sets, SEACR peaks found in 4 out of 5 NPAS4 replicates (2 h KA stimulation), 3 out of 3 NPAS4 replicates (0 h KA stimulation), 2 out of 2 ARNT2 replicates (2 h KA stimulation), 2 out of 2 EP400 replicates (0 and 2 h KA stimulation; Abcam antibody), 2 out of 2 MRE11 replicates (0 and 2 h KA stimulation), 2 out of 2 ETL4 replicates (0 h KA stimulation) and 3 out of 4 ETL4 replicates (2 h KA stimulation) were intersected using bedtools/2.27.1 intersect bed. Peaks within 150 bp were merged. Finally, the maxima of CUT&RUN signal within 100 bp windows for each peak was calculated from spike-in normalized bigWig files using custom scripts. For FOS, ARNT2, EP400, ETL4 and MRE11, final peak calls were extended 200 bp upstream and downstream from this peak maxima to generate 500 bp peak calls for each factor and time point. Mm10 blacklisted regions[60] were filtered out using the following command: bedops/2.4.30 –not-element-of 1 [BLACKLIST_BED]. NPAS4 peaks were extended to 1 kb, as we found maximal enrichment for the Ebox (CAGATG) motif and bHLH–PAS motif (CGTG) in 1 kb regions extended from the peak maxima. During the revision process, an additional three replicates of NPAS4 CUT&RUN were added as confirmatory of the initial five replicates performed. These additional NPAS4 CUT&RUN replicates are available as bigWig files and raw FASTQ files in the GEO submission but were not used in the analysis of NPAS4 peak calling.

Peak annotations and motif finding

Peak annotations as enhancers, promoters or other were determined using the homer (v.4.9) function annotatePeaks.pl. NPAS4 peak annotations were performed on regions extending 1 kb from the peak maxima, as these regions were most enriched for the NPAS4 motif. FOS peak annotations were performed on regions extending 500 bp from the peak maxima. Active regulatory elements were defined as the union of reproducible ATAC-seq (in 3 out of 3 replicates) and H3K27ac peaks (in 3 out of 3 replicates) extending 500 bp from the maximal ATAC-seq signal in that regulatory element. Enhancers were defined as the union of intergenic and intronic binding site annotations. To find enriched motifs, the sequences underlying each peak were extracted using bedtools/2.27.1 getfasta command.

Sequences of equal length (1,000 bp for NPAS4 and MRE11, 500 bp for sBLISS-seq) were processed using Meme-ChIP (https://meme-suite.org/meme/tools/meme-chip)[72] and tested against the motif background (HOCOMOCO v11) with significance reported as the *E*-value.

Peak overlaps

To determine the overlap between NPAS4 and additional factors, NPAS4 summits were extended to 1 kb. Peak overlaps were determined using bedtools/2.27.1 intersect bed allowing a minimum of 1 bp overlap. To generate sites with nonoverlapping signal for NPAS4 and FOS, NPAS4 and FOS summits were both extended to 1 kb. NPAS4-bound sites (no FOS) were generated using bedtools/2.27.1 intersect bed -v option to generate peaks with no NPAS4 and vice versa.

Generation of fixed line plots and aggregate plots

Fixed line plots were generated using homer(v4.9)'s annotatePeaks.pl [PEAK_BED] mm10 -d [INPUT_TAG_DIRS] -size 2000 –ghist -hist 25 -noann -nogene. Fixed line plots were generated from tag directories containing merged bam information from all replicates. Aggregate plots were generated using homer(v4.9)'s annotatePeaks.pl function with default parameter -hist 25, unless otherwise noted in the legend. Signal intensities were plotted using custom R scripts R (3.6.1). Aggregate plots show the average signal across replicates with the s.e.m. plotted. For statistical analysis of aggregate plots, signals for all replicates were extracted across the window specified in the figure legends using homer(v4.9)'s annotatePeaks.pl [PEAK_BED] mm10 -d [INPUT_TAG_DIRS] with homer's default read depth normalization to 10 million reads. Signals were averaged across all replicates, and Wilcoxon rank-sum tests were used to compare average signals between different conditions across the specified windows.

Identification of regulatory landscape in hippocampal neurons

To define a set of regulatory elements across hippocampal neuronal tissue samples, we used ATAC-seq (in *CamkIIa*$^+$ hippocampal neurons), together with CUT&RUN for the active histone modification H3K27ac, to characterize the constitutive and activity-responsive genomic regulatory element landscape in the hippocampus. We profiled hippocampal tissues in the basal state and 2 h following the synchronous induction of neuronal activity by low-dose KA administration. Reproducible ATAC-seq peaks were defined as MACS2 peaks identified in all 3 out of 3 replicates per time point. Sites from unstimulated and stimulated samples were concatenated and merged to generate a list of all possible ATAC-seq peaks across any stimulation condition. Reproducible H3K27ac peaks were defined as SEACR peaks identified in all 3 out of 3 replicates per time point. Sites from unstimulated and stimulated samples were concatenated and merged to generate a list of all possible H3K27ac peaks across any stimulation condition. The final landscape of elements included 179,841 elements, defined as the union of reproducible ATAC-seq and H3K27ac peaks found in our CUT&RUN and ATAC-seq datasets across all time points after removing mm10 blacklist regions. Note that blacklist removal was performed on the union of H3K27ac and ATAC-seq sites. To generate regions of comparable size, we determined the maximal ATAC-seq signal in each regulatory element and extended the regions 500 bp from this maximal chromatin accessibility summit.

To identify inducible ATAC-seq and H3K27ac peaks, we conducted a differential expression analysis using DeSeq2 (v.DESeq2_1.26.0)[59] on regions that had non-zero counts in at least two of the samples. We defined 'inducible elements' within this set as those sites exhibiting an increased ATAC signal (two-fold increase; adjusted $P < 0.05$ DeSeq2; P values were determined by testing against a fold change threshold of 2) and/or an increased H3K27ac CUT&RUN signal (1.5-fold increase; adjusted $P < 0.05$ DeSeq2; P values were determined by testing against a fold change threshold of 1.5) following stimulation (11,114 sites). Regulatory elements are provided in Supplementary Table 4. Note that elements that did not meet the threshold cut-off of non-zero counts in at least two of the samples will not be included in the respective DeSeq2 analysis tab in this table.

Quantification of CUT&RUN and sBLISS-seq signals at gene regulatory elements

To quantify transcription factor binding strength, the software package homer(v4.9) was first used to create tag directories for all replicates per factor for a given time point with the command makeTagDirectory [INPUT BED OR BAM]. For sBLISS-seq samples, bed files generated from a previously described pipeline[30] were expanded such that each UMI at each location was individually counted. For sBLISS-seq samples, the –len 0 tag was added to create tag directories to account for the fact that sBLISS-seq mapping generates bed files with single base-pair cut sites. Signals were extracted over sets of regulatory regions using the following command: annotatePeaks.pl [bed file] mm10 -size given -noann. For transcription factors (NPAS4, MRE11, ETL4, EP400, FOS and RAD50), signals were read-depth-normalized using homer(v4.9)'s default normalization to 10 million reads. To plot signals of various genomic features as a function of NPAS4 binding strength, we ranked all regulatory elements in our dataset (see above for the definition of 179,841 ATAC/H3K27ac elements in the hippocampus) according to NPAS4 IgG-normalized CUT&RUN signals and split regions into quartiles based on this ranking. We determined the NPAS4-normalized signal on a per site basis by dividing the aggregate signal for NPAS4 CUT&RUN (merge of NPAS4 replicates 1–5) by the aggregate signal for IgG (merged replicates A–F) at any given site in our regulatory landscape. NPAS4 Q4 (high) sites were determined as sites that overlapped with a SEACR-defined peak of NPAS4 and were in the top quartile of the NPAS4 IgG-normalized CUT&RUN signal. NPAS4 Q1 (low) sites had no SEACR-defined peak and were in the lowest quartile of the NPAS4 IgG-normalized CUT&RUN signal.

To compare γH2AX, ATAC-seq, H3K27ac, MRE11 and sBLISS-seq signals within the set of 179,841 defined in our regulatory landscape across time points and conditions (for example, different genotypes), raw read counts were extracted using homer(v4.9)'s annotatePeaks.pl function with the following command: annotatePeaks.pl [bed file] mm10 -size given -noann –noadj. For sBLISS-seq signals, the –len 0 parameter was included to quantify only the cut site end (1 bp) and to prevent read shifting based on estimated fragment sizes. Raw signal counts were imported in R. Before

running DeSeq2 (v.1.26.0), regions with low counts were excluded as follows: for ATAC-seq, H3K27ac and γH2AX, we required non-zero counts in at least two samples. For sBLISS-seq (both wild-type time course and Cre versus ΔCre datasets), which is sparser than ATAC-seq, H3K27ac CUT&RUN and γH2AX, we only eliminated regions with zero counts across all samples. DeSeq2-normalized counts were generated using the DeSeq2 function counts(dds, normalized=TRUE)[59] for our regulatory region set across all conditions (0 versus 2 h KA for ATAC, H3K27ac and γH2AX and 0, 2, 10 h KA for sBLISS-seq). We observed a batch processing effect for our numerous wild-type time course samples. Samples coming from all time points were included in each batch. However, to remove these effects computationally, we included the batch in our DeSeq2 design with the following command: DESeqDataSetFromMatrix(design ~Seq_Batch + Condition). In addition, when exporting normalized counts, we used the limma(3.42.2)limma::removeBatchEffect(counts(dds object, normalized=TRUE). Batches for the processing samples can be found in Supplementary Table 2. Signal counts plotted in boxplots in the paper for ATAC-seq, H3K27ac, γH2AX and wild-type sBLISS-seq time course are based on DeSeq2-normalized values averaged across all replicates for a given peak.

For processing the of Cre and ΔCre datasets, DeSeq2 normalization of counts was performed within each genotype (that is, raw counts across Cre versus ΔCre datasets in *Npas4*$^{fl/fl}$ were independently normalized from Cre versus ΔCre in wild-type datasets). We chose this design because these experiments were performed at separate times and comparisons were made between Cre and ΔCre within each genotype rather than across genotypes. To account for variability in sBLISS-seq datasets generated from very low cell numbers in the Cre versus ΔCre datasets, replicates consisting of independently infected animals were not averaged on a per peak basis but rather all replicate information was retained in plotting signals (Fig. 5b and Extended Data Fig. 12a–d).

Quantification of unique breaks across the genome

To compare breaks across the genome, input reads were downsampled to the lowest number of reads for all samples compared (NPAS4 Cre versus

ΔCre and wild-type Cre versus ΔCre: 21,036,891 reads; TIP60 Cre versus ΔCre 14,262,877 reads). We then quantified the total number of UMIs in each sample, which is plotted in Fig. 5c and Extended Data Fig. 12e. Note that for our quantification of signals across different regulatory elements displayed in all other figures, which used DESeq2-normalized counts, we did not downsample inputs before running DESeq2. The DESeq2 algorithm accounts for sequencing depth differences in normalization pipeline.

Quantification of mutational accumulation

Debarcer output (bamPositionComposition) tables were obtained from the package Debarcer_v.0.3.1 (https://github.com/oicr-gsi/debarcer/releases/tag/v0.3.1)[38]. These tables document the number of consensus ten families (that is, groups of ten reads with the same UMI in which all ten reads show the same base changes (or lack thereof)) for each base in a given amplicon. These tables also calculate the total number of UMI families that either match the reference genome or show a change from the reference. To include a given amplicon in our analysis, we required that the average number of consensus ten families across all bases in the amplicons was >100. For information on primer pooling and the amplicons included in the final analysis, see Supplementary Table 5.

Using scripts in R, we calculated the mutation frequency for a given amplicon by totalling the sum of all base changes from the reference and dividing by the total number of bases assessed in the amplicon (that is, the sum of consensus depth ten families across all bases in our table). The first 22 bases of the sequence, which contains the regions at which the primer anneals to amplify the region, was excluded from analysis to avoid errors in primer production complicating results. This calculation gives a single mutation rate for a given amplicon in a given sample. The total mutation rate includes both insertions and deletions and single nucleotide changes. To calculate the frequency of select point mutations, we counted the total number of select changes (C>A/(G>T)) divided by the total number of the given base included in the amplicon. Because it is not possible to know on which strand a mutation occurred, complementary base changes were collapsed into a single category (for example, C>T was combined with G>A). Insertion and deletion frequency was also calculated as a separate

category. We also calculated a per amplicon normalized mutation rate in which we divided the total mutation rate for each animal by the median total mutation frequency in young mice. For ageing gradient samples, wild-type mice aged 3 months old were considered young, 12 months old were considered middle aged and 23–27 months old were considered old. Extreme outlier points of both normalized mutation frequency and non-normalized frequency were removed across all samples using a ROUT's test at 0.1% confidence (Fig. 5e and Extended Data Fig. 13g). Outlier removal and statistical tests on mutational samples were performed in Prism (v.8.4.2).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequencing data have been deposited into the GEO with accession number GSE175965. Mass spectrometry data have been deposited to the ProteomeXchange Consortium through the PRIDE database with accession number PXD038718. Raw gel images are provided in Supplementary Fig. 1. Source data are provided with this paper.

Code availability

Custom code used in this study is available upon request.

References

1. Yap, E. L. & Greenberg, M. E. Activity-regulated transcription: bridging the gap between neural activity and behavior. *Neuron* **100**, 330–348 (2018).

2. Suberbielle, E. et al. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-β. *Nat.*

Neurosci. **16**, 613–621 (2013).

3. Madabhushi, R. et al. Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell* **161**, 1592–1605 (2015).

4. Delint-Ramirez, I. et al. Calcineurin dephosphorylates topoisomerase IIβ and regulates the formation of neuronal-activity-induced DNA breaks. *Mol. Cell* **82**, 3794–3809.e8 (2022).

5. Crowe, S. L., Movsesyan, V. A., Jorgensen, T. J. & Kondratyev, A. Rapid phosphorylation of histone H2A.X following ionotropic glutamate receptor activation. *Eur. J. Neurosci.* **23**, 2351–2361 (2006).

6. Yap, E. L. et al. Bidirectional perisomatic inhibitory plasticity of a Fos neuronal network. *Nature* **590**, 115–121 (2021).

7. Bloodgood, B. L., Sharma, N., Browne, H. A., Trepman, A. Z. & Greenberg, M. E. The activity-dependent transcription factor NPAS4 regulates domain-specific inhibition. *Nature* **503**, 121–125 (2013).

8. Lu, T. et al. Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).

9. Bunch, H. et al. Transcriptional elongation requires DNA break-induced signalling. *Nat. Commun.* **6**, 10191 (2015).

10. Ju, B. G. et al. A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* **312**, 1798–1802 (2006).

11. Iyama, T. & Wilson, D. M. 3rd DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst.)* **12**, 620–636 (2013).

12. Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).

13. Schumacher, B., Pothof, J., Vijg, J. & Hoeijmakers, J. H. J. The central role of DNA damage in the ageing process. *Nature* **592**, 695–703

(2021).

14. Lin, Y. et al. Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature* **455**, 1198–1204 (2008).

15. Doyon, Y. & Cote, J. The highly conserved and multifunctional NuA4 HAT complex. *Curr. Opin. Genet. Dev.* **14**, 147–154 (2004).

16. Galarneau, L. et al. Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. *Mol. Cell* **5**, 927–937 (2000).

17. Allard, S. et al. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J.* **18**, 5108–5119 (1999).

18. Allen Brain Map, Cell Types Database: RNA-Seq Data. https://celltypes.brain-map.org/rnaseq/human_m1_10x (2020).

19. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).

20. Pradhan, S. K. et al. EP400 deposits H3.3 into promoters and enhancers during gene activation. *Mol. Cell* **61**, 27–38 (2016).

21. Bird, A. W. et al. Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* **419**, 411–415 (2002).

22. Ikura, T. et al. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102**, 463–473 (2000).

23. Sun, Y. et al. Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. *Nat. Cell Biol.* **11**, 1376–1382 (2009).

24. Kusch, T. et al. Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science* **306**, 2084–2087 (2004).

25. Tang, J. et al. Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. *Nat. Struct. Mol. Biol.* **20**, 317–325 (2013).

26. Jacquet, K. et al. The TIP60 complex regulates bivalent chromatin recognition by 53BP1 through direct H4K20me binding and H2AK15 acetylation. *Mol. Cell* **62**, 409–421 (2016).

27. Brigidi, G. S. et al. Genomic decoding of neuronal depolarization by stimulus-specific NPAS4 heterodimers. *Cell* **179**, 373–391.e27 (2019).

28. Sharma, N. et al. ARNT2 tunes activity-dependent gene expression through NCoR2-nediated repression and NPAS4-mediated activation. *Neuron* **102**, 390–406.e9 (2019).

29. Dellino, G. I. et al. Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nat. Genet.* **51**, 1011–1023 (2019).

30. Bouwman, B. A. M. et al. Genome-wide detection of DNA double-strand breaks by in-suspension BLISS. *Nat. Protoc.* **15**, 3894–3941 (2020).

31. Canela, A. et al. Genome organization drives chromosome fragility. *Cell* **170**, 507–521.e18 (2017).

32. Wu, W. et al. Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).

33. Paull, T. T. 20 years of Mre11 biology: no end in sight. *Mol. Cell* **71**, 419–427 (2018).

34. Hoa, N. N. et al. Mre11 is essential for the removal of lethal topoisomerase 2 covalent cleavage complexes. *Mol. Cell* **64**, 1010 (2016).

35. Wienert, B. et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).

36. Salifou, K. et al. Chromatin-associated MRN complex protects highly transcribing genes from genomic instability. *Sci. Adv.* **7**, eabb2947 (2021).

37. Loonstra, A. et al. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl Acad. Sci. USA* **98**, 9209–9214 (2001).

38. Stahlberg, A. et al. Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. *Nat. Protoc.* **12**, 664–682 (2017).

39. Ooe, N., Motonaga, K., Kobayashi, K., Saito, K. & Kaneko, H. Functional characterization of basic helix-loop-helix-PAS type transcription factor NXF in vivo: putative involvement in an "on demand" neuroprotection system. *J. Biol. Chem.* **284**, 1057–1063 (2009).

40. Reid, D. A. et al. Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons. *Science* **372**, 91–94 (2021).

41. Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazzio, T. G. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* **22**, 999–1007 (2015).

42. Shen, Y. et al. RNA-driven genetic changes in bacteria and in human cells. *Mutat. Res.* **717**, 91–98 (2011).

43. Keskin, H. et al. Transcript-RNA-templated DNA recombination and repair. *Nature* **515**, 436–439 (2014).

44. Chahrour, M. H. et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet.* **8**, e1002635 (2012).

45. Humbert, J. et al. De novo KAT5 variants cause a syndrome with recognizable facial dysmorphisms, cerebellar atrophy, sleep disturbance, and epilepsy. *Am. J. Hum. Genet.* **107**, 564–574 (2020).

46. Bell, S. et al. Mutations in *ACTL6B* cause neurodevelopmental deficits and epilepsy and lead to loss of dendrites in human neurons. *Am. J. Hum. Genet.* **104**, 815–834 (2019).

47. Tian, X. et al. SIRT6 is responsible for more efficient DNA double-strand break repair in long-lived species. *Cell* **177**, 622–638.e22 (2019).

48. Zullo, J. M. et al. Regulation of lifespan by neural excitation and REST. *Nature* **574**, 359–364 (2019).

49. Fisher, J. B. et al. Depletion of Tip60 from in vivo cardiomyocytes increases myocyte density, followed by cardiac dysfunction, myocyte fallout and lethality. *PLoS ONE* **11**, e0164855 (2016).

50. Chen, P. B. et al. Hdac6 regulates Tip60-p400 function in stem cells. *eLife* **2**, e01557 (2013).

51. Buis, J., Stoneham, T., Spehalski, E. & Ferguson, D. O. Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2. *Nat. Struct. Mol. Biol.* **19**, 246–252 (2012).

52. Buis, J. et al. Mre11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation. *Cell* **135**, 85–96 (2008).

53. Mo, A. et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).

54. Deisseroth, K., Bito, H. & Tsien, R. W. Signaling from synapse to nucleus: postsynaptic CREB phosphorylation during multiple forms of hippocampal synaptic plasticity. *Neuron* **16**, 89–101 (1996).

55. Erbel-Sieler, C. et al. Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc. Natl Acad. Sci. USA* **101**, 13648–13653 (2004).

56. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

57. Hainer, S. J. & Fazzio, T. G. High-resolution chromatin profiling using CUT&RUN. *Curr. Protoc. Mol. Biol.* **126**, e85 (2019).

58. Wong, N., John, S., Nussenzweig, A. & Canela, A. END-seq: an unbiased, high-resolution, and genome-wide approach to map DNA double-strand breaks and resection in human cells. *Methods Mol. Biol.* **2153**, 9–31 (2021).

59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

60. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

61. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

62. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

63. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).

64. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

65. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2—an R package for gene list functional enrichment analysis

and namespace conversion toolset g:Profiler. *F1000Res* https://doi.org/10.12688/f1000research.24956.2 (2020).

66. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

67. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

68. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Systems* **8**, 329–337.e4 (2019).

69. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).

70. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

71. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).

72. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

Acknowledgements

We thank all Greenberg Laboratory members for valuable input; S. Hinshaw for advice and assistance with biochemistry experiments; J. Lough for *Tip60*$^{fl/fl}$ mice; T. Fazzio for *Tip60-H3F* mice; D. Ferguson for *Mre11*$^{fl/fl}$ mice; N. Crosetto and B. Bouwman for access to the sBLISS-seq protocols; T. Godfrey for advice on amplicon library production; T. Vierbuchen, J. Green, J. Tycko and A. Greben for manuscript comments and helpful insight throughout the course of the project; and J. Luquette for advice and

assistance on mutational analysis. E.A.P. acknowledges a Good Ventures Life Sciences Research Fellowship and K99 from NIA 1K99AG064042-01A1. D.T.G. was supported by a Harvard Department of Neurobiology Graduate fellowship. C.P.D. was supported by NIH fellowships T32-NS007473 and F32-NS112455. E.-L.Y. was supported by a Stuart H. Q. and Victoria Quan fellowship, a Harvard Department of Neurobiology Graduate fellowship and an Aramont Fund for Emerging Science Research fellowship. M.A.N. acknowledges NIH fellowship T32GM007753. E.E.D. was supported by the Damon Runyon Cancer Research Foundation and a Warren Alpert Distinguished Scholar Fellowship Award. Histology imaging was performed through the Harvard Medical School Neuro Imaging Facility (NINDS P30 Core Center grant number NS072030). This work was supported by R01 NS028829, the Lefler Faculty Small Grant to M.E.G. and the Carol and Gene Ludwig Family Foundation through the Ludwig Neurodegenerative Disease Seed Grants Program at Harvard Medical School. The Greenberg Laboratory is supported by the Allen Discovery Center Program, a Paul G. Allen Frontiers Group advised programme of the Paul G. Allen Family Foundation and the Tang-Yang Autism Center at Harvard Medical School.

Author information

Author notes

1. Elizabeth A. Pollina

   Present address: Department of Developmental Biology, Washington University School of Medicine, St Louis, MO, USA

2. These authors contributed equally: Elizabeth A. Pollina, Daniel T. Gilliam

Authors and Affiliations

1. Department of Neurobiology, Harvard Medical School, Boston, MA, USA

Elizabeth A. Pollina, Daniel T. Gilliam, Cindy Lin, Naomi Pajarillo, Christopher P. Davis, David A. Harmin, Ee-Lynn Yap, Ian R. Vogel, Eric C. Griffith, M. Aurel Nagy, Emi Ling, Erin E. Duffy, Charles J. Weitz & Michael E. Greenberg

2. Department of Neurobiology, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA

Andrew T. Landau & Bernardo L. Sabatini

Contributions

E.A.P., D.T.G. and M.E.G. conceived and designed the study and wrote the manuscript. E.A.P. and D.T.G. designed, executed and analysed the majority of experiments performed in the paper. Specifically, E.A.P. designed and characterized the *Npas4–FH* and *Arnt2–FH* lines, performed and analysed biochemistry, mass spectrometry, RNA-seq, ATAC-seq, CUT&RUN, ChIP–seq, sBLISS-seq, lifespan, FACS sorting, amplicon and mutation library preparation and analyses. D.T.G. performed and analysed biochemistry, mass spectrometry, RNA-seq, snRNA-seq, sBLISS-seq, END-seq, amplicon and mutation library production, all stereotaxic surgery and assisted in electrophysiology experiments. A.T.L. performed electrophysiology experiments. C.L. performed biochemical assays, RNA-seq and assisted in amplicon library production. N.P. performed biochemical assays, RNA-seq and assisted in amplicon library production. D.A.H. performed additional mutational analysis. E.-L.Y. and C.P.D. performed FOS and H3K27ac CUT&RUN experiments. C.P.D. provided code for CUT&RUN data analyses. I.R.V. assisted with sBLISS-seq library production. M.A.N. assisted in the design and production of the *Npas4–FH* and *Arnt2–FH* tagged lines. E.L. performed FOS and EGR1 immunoprecipitation–mass spectrometry experiments. E.E.D. assisted in GO analysis of snRNA-seq data. E.C.G. contributed to project direction and manuscript writing. C.J.W. provided intellectual guidance and reagents for biochemical purification of NPAS4 complexes. B.L.S. provided intellectual guidance for electrophysiology experiments.

Corresponding author

Correspondence to Michael E. Greenberg.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Jan Hoeijmakers, Eran Mukamel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher′s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 NPAS4 forms a complex with NuA4 across brain regions.

a. Lysate from the hippocampus of KA-stimulated mice, fractionated by molecular weight (MW) using gel filtration and non-denaturing size exclusion chromatography. Western blotting for NPAS4 in the different MW fractions confirms that NPAS4 resides in a high MW complex peaking ~1 MDa in size. Representative image from 3 experiments. See gel source data (Supplementary Fig. 1). b. Sequence validation of Flag HA epitope tag appended to the C-terminus of *Npas4* or *Arnt2* in the P0 or F1 generation in *Npas4-FH* or *Arnt2-FH* mice. c. Left: Immunohistochemistry of NPAS4 and HA antibody staining in hippocampus of *Npas4-FH* mice, 6 h enriched

environment. Scale bars: 100 µm. Right: Validation of *Arnt2-FH* knockin mouse line by immunohistochemistry in CA1. Scale bar 25 µm. Representative images from 3 experiments. d. Validation by Western blot that NPAS4 and NPAS4-FH have the same induction kinetics and specificity for membrane depolarization-induced calcium signaling. To stimulate cultured cortical neurons, 55 mM KCl was applied for the indicated time points. Representative image from one experiment. See gel source data (Supplementary Fig. 1). e. Confirmation by Western blot of NPAS4-FH expression and Flag immunoprecipitation (IP) in cultured cortical neurons from either wild-type or *Npas4-FH* mice. Representative image from one experiment. See gel source data (Supplementary Fig. 1). f. Confirmation via anti-Flag immunoprecipitation (IP) and Western blot that NPAS4-NuA4 also assembles in the visual cortex following 2 h light stimulation. Representative image from 2 experiments. See gel source data (Supplementary Fig. 1). g. Experimental diagram detailing glycerol gradient fractionation and immunoprecipitation of intact NPAS4-NuA4 complexes from high molecular weight (MW) fractions. h. Western blot of glycerol gradient fractions of NPAS4-FH in mouse cortical lysates segregated by MW. NPAS4 migrates in high MW fractions along with NuA4 components TRRAP, EP400, and DMAP1. Representative image from 3 experiments. See gel source data (Supplementary Fig. 1). i. Reciprocal validation of NPAS4-NuA4 interaction via Flag immunoprecipitation and Western blotting from high MW fractions of cortical lysates from wild-type, *Npas4-Flag-HA (Npas4-FH)*, and *Tip60-Flag (Tip60-F)* mice. Representative image from 2 experiments. See gel source data (Supplementary Fig. 1). j. Reciprocal validation of NPAS4-NuA4 interaction via Flag immunoprecipitation and Western blotting from low MW fractions of cortical lysates from wild-type, *Npas4-Flag-HA (Npas4-FH)*, and *Tip60-Flag (Tip60-F)* mice. Representative image from 2 experiments. See gel source data (Supplementary Fig. 1).

Extended Data Fig. 2 Validation of NPAS4-NuA4 interaction and neuronal specificity.

a. Confirmation that the NPAS4-NuA4 interaction can be observed via reciprocal Flag IP and Western blot from hippocampi of wild-type, *Npas4-FH*, and *Tip60-Flag (TIP60-F)* mice. Representative image from 2

experiments. See gel source data (Supplementary Fig. 1). b. Anti-Flag IP from hippocampi of *Tip60-Flag (TIP60-F)* mice in both unstimulated brains and 2 h post KA stimulation. Western blot demonstrates that components of the NuA4 complex interact in both the basal (0 h) and 2 h condition while NPAS4 interacts primarily in the stimulated state. Representative image from 2 experiments. See gel source data (Supplementary Fig. 1). c. Average RNA-seq DeSeq2 normalized counts ± s.e.m. of NPAS4-NuA4 components in the mouse hippocampus at 0 h (n = 8), 2 h (n = 8), and 10 h (n = 10), post KA stimulation. ***P* = 3.19e-29. *P* value determined from transcriptome-wide DeSeq2 analysis with Benjamini-Hochberg correction. See source data for individual *P* values. d. Full length NPAS4 (amino acids 1-802) and the indicated truncations of the C-terminal portion of NPAS4 were expressed in HEK293T cells. Sequential Flag and HA IPs were performed (right), followed by Western blotting for NPAS4, ARNT2, and NuA4 subunits TRRAP and DMAP1. Flag-HA-tagged nuclear GFP (nGFP) was included as negative control. Representative image from 3 experiments. See gel source data (Supplementary Fig. 1). e. Heatmap of the column-normalized specificity score for each component of the NuA4 complex in anti-Flag IP-MS experiments conducted with NPAS4/ARNT, NPAS4/ARNT2, FOS/JUN, or EGR1 expressed in HEK293T cells. All proteins were Flag-HA tagged. Specificity score represents the ratio of the number of peptides identified in the IP over the number of peptides found in nGFP controls performed in parallel. Replicate numbers provided in Supplementary Table 1. f. Normalized counts (Seurat v3) of NPAS4-NuA4 components from hippocampal single-nucleus RNA sequencing in mouse hippocampus, normalized across column and displayed as a Z-score. g. Marker genes identifying cell types in human primary motor cortex single-nucleus RNA sequencing dataset published by the Allen Brain Institute (see Fig. 1d)[18], normalized across column and displayed as a Z-score.

Source Data

Extended Data Fig. 3 NPAS4 and NuA4 co-localize on chromatin across the genome.

a. Aggregate plot of average CUT&RUN coverage (fragment depth per bp/per peak) ± s.e.m. at NPAS4-binding sites in *Npas4* wild-type vs *Npas4*

KO hippocampal tissue. (*Npas4* wild-type: n = 5, *Npas4* KO: n = 2, IgG: n = 8) 3-5 mice pooled per replicate. ***$P < 2.2e-16$, P values were on calculated on average signal extracted in a 2 kb window centered on NPAS4 peak summits using unpaired, two-tailed Wilcoxon rank-sum tests. b. Integrative Genomics Viewer tracks of aggregated NPAS4 CUT&RUN signal (n = 5). 3-5 mice pooled per replicate. 3 additional replicates of NPAS4 CUT&RUN signal confirm the reproducibility of NPAS4 CUT&RUN. c. Distribution of NPAS4 (10,225 sites) and FOS (11,770 sites) CUT&RUN peak annotations relative to all regulatory elements (179,841 sites) in hippocampal tissue. d. Significant motifs enriched in 10,225 NPAS4 CUT&RUN peaks. E-value calculated using MEME-ChIP[72]. e. Correlation between NPAS4 CUT&RUN and NPAS4 ChIP-seq reads at NPAS4 ChIP-seq peaks (10,917 ChIP-seq peaks). Values represent the $\log_2$ read depth normalized counts for CUT&RUN vs ChIP-seq. Correlation was calculated by Pearson (R = 0.45) and Spearman (Rho = 0.41) tests, $P < 2.2e-16$ by two-tailed correlation tests. The reciprocal analysis of the correlation between NPAS4 CUT&RUN and NPAS4 ChIP-seq reads at NPAS4 CUT&RUN peaks (10,225 CUT&RUN peaks) yields a Pearson R = 0.46 and Spearman Rho = 0.43, $P < 2.2e-16$ by two-tailed correlation tests. f. Venn diagram of overlaps between SEACR peaks for anti-EP400 Ab1 (antibody 1, Bethyl Labs, A300-541A) and anti-EP400 Ab2 (antibody 2, Abcam Ab5201). Peak overlaps indicate at least 1 bp overlap between the entire SEACR-enriched region. Maxima indicates overlap between regions extended 500 bp out from the peak maxima for each factor. g. Venn diagram of overlaps between SEACR peaks for NPAS4/ARNT2 co-bound peaks with ETL4 and high-confidence EP400 peaks, defined as the union of peaks shared by both EP400 antibodies. Maxima indicates overlap between regions extended 500 bp out from the peak maxima for each factor. h. Boxplot of average EP400 normalized signal (counts per million) at sites bound by FOS but not NPAS4 (labeled 'FOS') and sites bound by NPAS4 but not FOS (labeled 'NPAS4'). FOS/No NPAS4 sites are defined as sites with a SEACR-determined peak of CUT&RUN signal for FOS but no peak for NPAS4 (6,998 sites) and vice versa (NPAS4/No FOS peaks: 5,550 sites). Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). 3-5 mice pooled per replicate. Replicate numbers provided in Supplementary Table 2. ***$P < 2.2e-16$. P values were calculated using

unpaired, two-tailed Wilcoxon rank-sum tests. i. Representative Integrative Genomics Viewer tracks of replicate EP400 CUT&RUN at the *Bdnf* promoter.

Source Data

Extended Data Fig. 4 Acute deletion of NPAS4 or TIP60 by viral injection and additional quality control for single-nucleus RNA-seq datasets.

a. Immunohistochemistry image of an *Npas4*$^{fl/fl}$ mouse injected with AAV to express Cre-mCherry (shown in red) and collected 2 h post low-dose KA to induce NPAS4 (shown in cyan). Representative image from 3 animals. Scale bar = 1 mm. b. Immunohistochemistry image of a *Tip60*$^{fl/fl}$ mouse injected with AAV to express Cre-mCherry (shown in red). TIP60 (shown in cyan). Representative image from 3 animals. Scale bar = 1 mm. c. Immunohistochemistry image of both hippocampal hemispheres of an *Npas4*$^{fl/fl}$ mouse injected with Cre-mCherry and ΔCre-GFP in contralateral sides of the hippocampus and collected 2 h post low-dose KA to induce NPAS4 (shown in white). Representative image from 3 animals. Scale bar = 1 mm. d. Western blot from whole hippocampal tissue of *Tip60*$^{fl/fl}$ mice injected with AAV expressing Cre-mCherry and ΔCre-GFP in contralateral sides of the hippocampus. Tissue was collected at either 0 or 2 h post KA stimulation. Cre and ΔCre tissue was collected from each individual mouse (0 h, n = 2; 2 h, n = 2). See gel source data (Supplementary Fig. 1). e. Quantification of the Western blot is shown in d, normalizing the NPAS4 signal to loading control GAPDH. f. Left: UMAP visualization of full *Npas4*$^{fl/fl}$ snRNA-seq dataset. Nuclei are colored according to mouse of origin. 32,418 nuclei from 2 mice. Right: UMAP visualization of full *Tip60*$^{fl/fl}$ snRNA-seq dataset. Nuclei are colored according to mouse of origin. 44,511 nuclei from 3 mice. g. Summary of final nuclei numbers in *Npas4*$^{fl/fl}$ and *Tip60*$^{fl/fl}$ snRNA-seq datasets, and quantification of infection rates with Cre-mCherry and ΔCre-GFP viruses. Higher infection rate in neurons reflects the known tropism of the AAV2/9 virus used in these experiments. h. Cell-type assignment in *Npas4*$^{fl/fl}$ (left) and *Tip60*$^{fl/fl}$ (right) snRNA-seq datasets using indicated marker genes. The y-axis denotes

normalized expression (Seurat $\log_e$ normalized counts). i. Distribution of number of genes detected per nucleus, by cell-type. *Npas4$^{fl/fl}$* (left) and *Tip60$^{fl/fl}$* (right).

Source Data

Extended Data Fig. 5 NPAS4-NuA4 coordinate gene regulation across neuronal subtypes *in vivo*.

a. Heatmap showing coordinate regulation of NPAS4 target genes across the principal neuronal subtypes of the hippocampus in both *Npas4$^{fl/fl}$* and *Tip60$^{fl/fl}$* mice. NPAS4 target genes were identified in each cell-type using both Seurat (v3) and Monocle3 (see Methods). Bonferroni adjusted *P* values from Seurat (v3) differential expression testing (unpaired, two-tailed Wilcoxon rank-sum test) between ΔCre- and Cre-infected nuclei are shown in each cell-type. Each column represents adjusted *P* values for one gene. b. Violin plots showing the distribution of expression (Seurat $\log_e$ normalized counts) of the indicated gene across nuclei in the indicated cell-type. Bonferroni adjusted *P* value $***P < 2.2e-16$. Differential gene expression tests were conducted using an unpaired, two-tailed Wilcoxon rank-sum test implemented via the Seurat (v3) FindMarkers function. c. Comparison of the observed differences (ΔCre – Cre) in normalized counts for NPAS4 or TIP60 target genes in each neuronal subtype to the differences obtained when using an equal number of randomly selected genes. Genes were randomly selected from the top 10% of expressed genes (to account for NPAS4 or TIP60 target genes being highly expressed on average), and the average difference (ΔCre – Cre) in expression for each gene was calculated for each random sample. This sampling was repeated 10,000 times to generate sampling distributions (gray). In each subtype, the average difference (ΔCre – Cre) observed when using that subtype's NPAS4 or TIP60 target genes lies far outside the distribution obtained using randomly selected genes, suggesting the differences in expression of the target genes between ΔCre- and Cre-infected nuclei is not due to chance. d. Boxplots showing $\log_2$ fold change of NPAS4 or TIP60 target genes (see Methods) comparing ΔCre- and Cre-infected nuclei in dentate gyrus, CA1, CA3, and inhibitory neurons. Boxplot shows median (line), IQR (box), 1.5x

IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). *** $P < 2.2e-16$. P values were calculated using two-tailed, unpaired t-tests comparing to a null hypothesis of $\log_2$ fold change = 0 (no difference between Cre and ΔCre). Exact P values and cell numbers per cluster provided in source data. e. Select overlapping GO terms enriched in both NPAS4 target genes and TIP60 target genes across dentate gyrus, CA1, CA3, and inhibitory neurons. Circle size indicates the adjusted P value of enrichment determined by Fisher's one-tailed test using gProfiler2[65]. Color indicates the fold enrichment. See Methods for additional detail and Supplementary Table 3 for complete list of enriched GO terms for each cell-type.

Source Data

Extended Data Fig. 6 NPAS4-NuA4 coordinately regulates gene transcription and somatic inhibition.

a. Principal component analysis clustering of RNA-seq datasets from cultured mouse neurons expressing shRNAs targeting *Npas4*, *Tip60*, or *Ep400*, and stimulated with 55 mM KCl for either 0, 2, or 6 h. Control shRNA targets luciferase. b. Boxplots of $\log_2$ fold changes between the control shRNA (n = 3) and the indicated *Npas4* (n = 3), *Tip60* (n = 3) or *Ep400* (n = 3) shRNAs in neuronal cultures 6 h following membrane-depolarization by 55 mM KCl. Replicates consist of primary cultures generated on independent days. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). Activity-regulated-genes (ARGs) are defined as genes upregulated at least 1.5-fold with a Benjamini-Hochberg adjusted $P < 0.01$ comparing the 6 h and 0 h time points in control shRNA-treated neurons. *Tip60* and *Ep400* targets are defined as all genes down-regulated by at least 1.5-fold with a Benjamini-Hochberg adjusted $P < 0.01$ by both shRNAs. Non-targets include all expressed genes not significantly affected by loss of *Tip60* or *Ep400*. ***$P < 2.2e-16$. P values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests. c. Average expression ± s.e.m. of *Ep400* (n = 3), *Tip60* (n = 3), and *Npas4* (n = 3) by qPCR. Replicates consist of primary cultures generated on independent days. Expression normalized to both *Tubb3* and *Gapdh*. *Ep400* 0 h shRNA1: *$P = 0.0026$, shRNA2: *$P = 0.0072$;

2 h shRNA1: *$P = 0.0072$, shRNA2: *$P = 0.0154$; 6 h shRNA1: *$P = 0.0027$, shRNA2: *$P = 0.0038$. *Tip60* 0 h shRNA1: ns = 0.4021, shRNA2: ns = 0.2669; 2 h shRNA1: ns = 0.4507, shRNA2: ns = 0.3324; 6 h shRNA1: ns = 0.433, shRNA2: ns = 0.5431. *Npas4* 0 h shRNA1: *$P = 0.0214$; 2 h shRNA1: *$P = 0.0004$; 6 h shRNA1: *$P = 0.0001$. *P* values by two-tailed, unpaired t-tests. d. HA staining of cultured cortical neurons infected with nuclear GFP (nGFP) and Flag-HA-tagged NPAS4 (full length and with indicated truncations) (see Extended Data Fig. 2d). Representative image from one experiment. Scale bar = 50 µm. e. Luciferase activity of NPAS4-bound enhancers in cultured neurons transfected with either nGFP, full length NPAS4, or indicated NPAS4 truncations (see Extended Data Fig. 2d). Colors represent the different NPAS4 truncations. Average luciferase activity normalized to nGFP-expressing control samples ± s.e.m. (see Methods). Each point represents a value from an independently transfected well collected from at least 3 independent primary neuronal cultures, except for Peak1 Npas4_1-699 (2 cultures). **$P < 0.0045$; *$P = 0.049$. *P* values were calculated using two-tailed, unpaired t-tests with Benjamini-Hochberg correction for multiple hypothesis testing. Individual *P* values and replicate numbers provided in source data. f. Points represent distance between cells in an uninfected:Cre-infected cell pair. *Npas4$^{fl/fl}$*: Stim: n = 16 pairs from 4 mice, Unstim: n = 12 pairs from 5 mice. *Tip60$^{fl/fl}$*: Stim: n = 11 pairs from 2 mice, Unstim: n = 12 pairs from 2 mice. $P = 0.305$. *P* value was calculated using a one-way ANOVA. g. Lateral distance from center of each pair to the stimulating electrode placed in the center of stratum pyramidale. *Npas4$^{fl/fl}$*: Stim: n = 16 pairs from 4 mice, Unstim: n = 12 pairs from 5 mice. *Tip60$^{fl/fl}$*: Stim: n = 11 pairs from 2 mice, Unstim: n = 12 pairs from 2 mice. $P = 0.055$. *P* value was calculated using a one-way ANOVA. h,i. Access resistance in MΩ for each pair of simultaneously patched CA1 pyramidal neurons in (h) *Npas4$^{fl/fl}$*: Stim: n = 16 pairs from 4 mice, $P = 0.7396$; Unstim: n = 12 pairs from 5 mice, $P = 0.9677$ and (i) *Tip60$^{fl/fl}$*: Stim: n = 11 pairs from 2 mice, $P = 0.7533$; Unstim: n = 12 pairs from 2 mice, $P = 0.8906$. *P* values by unpaired, two-tailed t-tests.

Source Data

[Extended Data Fig. 7 NPAS4-NuA4 binds to regions undergoing recurrent DNA breaks *in vivo*.]

a. Definition of all regulatory elements used throughout the manuscript. Venn Diagram depicting the overlap of reproducible ATAC-seq peaks (merged between 0 and 2 h peaks) and reproducible H3K27ac CUT&RUN (merged between 0 and 2 h peaks). ATAC-seq and CUT&RUN peak sets were defined as peaks consistently found in 3 of 3 replicates per timepoint. 3-5 mice pooled per replicate. All regulatory elements are defined as the union of reproducible ATAC-seq and H3K27ac peaks across all timepoints. Activity-inducible regulatory elements are defined as elements that exhibit a greater than two-fold change in ATAC-seq signal (2 vs 0 h stimulation; adjusted $P < 0.05$) and/or a 1.5-fold change in H3K27ac CUT&RUN signal (2 vs 0 h stimulation; adjusted $P < 0.05$). P values calculated by DeSeq2's Wald test with default Benjamini-Hochberg correction. b. Upper Panel: Heatmap of Euclidean distance between replicates of ATAC-seq in hippocampal nuclei isolated from unstimulated (0 h) or 2 h post stimulation across all regulatory elements defined in Extended Data Fig. 7a. Lower panel: Representative Integrative Genomics Viewer tracks of individual ATAC-seq replicates at activity-inducible gene *Inhba*. c. Upper Panel: Heatmap of Euclidean distance between replicates of H3K27ac CUT&RUN in hippocampal nuclei isolated from unstimulated brains (0 h) or 2 h post stimulation across all regulatory elements defined in Extended Data Fig. 7a. Lower panel: Representative Integrative Genomics Viewer tracks of individual H3K27ac CUT&RUN replicates at activity-inducible gene *Bdnf*. d. Principal component analysis of γH2AX ChIP-seq signal across all regulatory elements (see Extended Data Fig. 7a) in unstimulated and stimulated hippocampal nuclei. Replicates cluster together and separate by stimulation state. e. Representative Integrative Genomics Viewer tracks of individual γH2AX ChIP-seq replicates and aggregate NPAS4 CUT&RUN signal (n = 5) at activity-inducible gene *Rgs7bp*. 3-5 mice pooled per replicate. f. Schematic of sBLISS-seq on cultured neurons infected with either Cas9-only viruses or Cas9 virus + gRNAs. g. Integrative Genome Browser tracks displaying sBLISS-seq signal at the *Fos* promoter in cultured neurons infected with either Cas9+gRNA to the *Fos* locus or a Cas9-only control. Red line indicates the position of the gRNA. Zoomed-in perspective shows the reads mapping on either side of the predicted gRNA

cut site, indicated by the arrow. PAM sites are underlined in the DNA sequence. Representative image from 3 experiments. Replicates consist of independent cultures generated on separate days. h. Integrative Genome Browser tracks displaying sBLISS-seq signal at the *Inhba* enhancer in cultured neurons infected with either Cas9+gRNA to the *Inhba* enhancer locus or a Cas9-only control. Representative image from 3 experiments. Replicates consist of independent cultures generated on separate days. i. Boxplots showing sBLISS-seq normalized signal (see Methods) across all regulatory elements in wild-type neurons at 0 (n = 8) and 2 h (n = 8) after KA stimulation. 1 mouse per replicate. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). Promoters = 11,853, Introns = 79,438, Intergenic = 81,385, 3′UTR = 1,574, 5′UTR = 695, non-coding = 445, TTS = 1,831, Exons = 2,620. See Extended Data Fig. 7a. j. Aggregate plot of average sBLISS-seq coverage (fragment depth per bp per peak) ± s.e.m. (n = 8) at the TSS of genes in the highest quartile of expression (Q4) vs genes in the lowest quartile (Q1) in 2 h stimulated hippocampi. Gene expression quartiles were determined from DeSeq2 normalized counts of paired RNA-seq samples collected from the same tissue as sBLISS-seq samples at the 2 h KA stimulation timepoint (see Fig. 3d). ***$P < 2.2e\text{-}16$. P value was calculated on average signal extracted in the 20 kb window around gene TSS using an unpaired, two-tailed Wilcoxon rank-sum test. k. Distribution of DSB peak annotations at 2 h of stimulation (4,447 peaks). Reproducible sBLISS-seq peaks were defined by the MACS2 peak calling algorithm and were found in at least 5 of 8 replicates. l. Correlation between sBLISS-seq signal (log_2 normalized counts) and γH2AX signal (log_2 normalized counts) across all regulatory elements in the hippocampus (179,841 sites; see Extended Data Fig. 7a). Correlation was calculated by Pearson (R = 0.47) and Spearman (Rho = 0.48), with $P < 2.2e\text{-}16$ for both two-tailed correlation tests. m. Most significant motifs enriched in reproducible sBLISS-seq peaks at 2 h of stimulation. Motifs of activity-inducible transcription factors (ATF1, EGR1, and NPAS4/AHR) are enriched in sBLISS-seq peaks. E-value calculated using MEME-ChIP[72]. n. Boxplots of average sBLISS-seq normalized counts at activity-inducible elements (11,114 sites) vs non-inducible elements (168,727 sites) at 0 h (n = 8 mice) and 2 h (n = 8 mice) post stimulation. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). Non-inducible

elements include all elements that do not fall within the inducible peak set in our regulatory landscape. ***$P = 2.2e\text{-}16$, ns: $P = 1$. P values by unpaired, one-tailed Wilcoxon rank sum test (2 h > 0 h).

Source Data

Extended Data Fig. 8 END-seq mapping of DSB signal in cultured cortical neurons shows enrichment at NPAS4-bound sites and activity-dependent dynamics.

a. Aggregate plots (bin size 50 bp) of average END-seq coverage (fragment depth per bp per peak) ± s.e.m. (n = 2) in cultured mouse cortical neurons at CTCF-bound sites (13,228 sites) at either 0 or 2 h of stimulation with 55mM KCl. See Methods for full description of END-seq method and experimental parameters. CTCF-bound sites are defined as sites with a SEACR-defined peak of CTCF CUT&RUN signal in hippocampal datasets. ***$P < 2.2e\text{-}16$. For a-c, replicates consist of independent primary cultures. P values by unpaired, two-tailed Wilcoxon rank-sum tests on signal extracted in a 500 bp window centered on each peak. b. Aggregate plots (bin size 50 bp) of average END-seq coverage ± s.e.m. (n = 2) in cultured mouse cortical neurons at NPAS4-bound sites (10,225 sites) at either 0 or 2 h post stimulation with 55mM KCl. NPAS4-bound sites are defined as sites with a SEACR-defined peak of NPAS4 CUT&RUN in *in vivo* hippocampal datasets. ***$P < 2.2e\text{-}16$. c. Aggregate plots (bin size 50 bp) of average END-seq coverage ± s.e.m. (n = 2) in cultured mouse cortical neurons at NPAS4 sites that lack a FOS peak (5,550 sites) and FOS sites that lack an NPAS4 peak (6,998 sites), at both 0 and 2 h post stimulation with 55mM KCl. ***$P < 2.2e\text{-}16$ for NPAS4 0 vs 2 h, ***$P = 2.3e\text{-}15$ for FOS 0 vs 2 h. d. Aggregate plots (bin size 50bp) of average sBLISS-seq coverage ± s.e.m. (n = 8) at CTCF-bound sites at 0 and 2 h post stimulation with low-dose KA. ***$P < 2.2e\text{-}16$. For d-h, replicates derived from individual mice. P values by unpaired, two-tailed Wilcoxon rank-sum tests on signal extracted in a 500 bp window centered on each peak. e. Aggregate plots (bin size 50 bp) of average sBLISS-seq coverage ± s.e.m. (n = 8) at NPAS4-bound sites at 0 and 2 h post stimulation. **$P < 2.2e\text{-}16$. f. Aggregate plots (bin size 50 bp) of average sBLISS-seq coverage ± s.e.m. (n = 8) at NPAS4 sites that lack a FOS peak and FOS sites that lack an NPAS4 peak at 2 h

post stimulation. ***$P < 2.2e-16$. g. Aggregate plots (bin size 50 bp) of average sBLISS-seq coverage ± s.e.m. (n = 8) at NPAS4 sites that lack a FOS peak at 0 and 2 h post stimulation. ***$P < 2.2e-16$. h. Aggregate plots (bin size 50 bp) of average sBLISS-seq coverage ± s.e.m. (n = 8) at FOS sites that lack an NPAS4 peak at 0 and 2 h post stimulation. ***$P < 2.2e-16$. i. Boxplots showing average NPAS4 (n = 5) and FOS (n = 3) CUT&RUN signal (counts per million) at NPAS4 sites that lack a FOS peak and FOS sites that lack an NPAS4 peak at 2 h post stimulation in hippocampus. 3-5 mice pooled per replicate. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). ***$P < 2.2e-16$. *P* values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests.

Extended Data Fig. 9 NPAS4-NuA4-bound sites undergo DNA repair as inducible transcription subsides.

a. Heatmap of Euclidean distance between replicates of RNA-seq prepared from the same tissue used to generate the sBLISS-seq timecourse. The samples cluster primarily according to stimulation state, with the 10 h post stimulation samples falling into two groups with either dampening or sustained transcriptional induction. b. DeSeq2 normalized counts of additional inducible genes, *Inhba* and *Cgref1*, are shown for each replicate. c. Principal component analysis clustering of sBLISS-seq samples at 0 h, 2 h and 10 h post stimulation. The sBLISS-seq samples cluster according to stimulation state, with the 10 h post stimulation samples clustering either with the 0 h or 2 h samples. Paired RNA-seq analysis indicates that this separation is driven by altered levels of transcriptional induction in these samples. d. Aggregate plots (bin size 50 bp) of average sBLISS-seq coverage (fragment depth per bp per peak) ± s.e.m. at all regulatory elements, subset by quartiles of NPAS4 CUT&RUN signal. Q1 = 44,864 sites, Q4 = 7,378 sites. 0 h (n = 8), 2 h (n = 8), 10 h (n = 10). 1 mouse per replicate. *P* values were calculated using unpaired, two-tailed Wilcoxon rank sum tests. NPAS4 Q1: 0vs2 h: $P < 2.2e-16$, 2 vs 10 h: $P = 1.7e-15$, 0 vs 10 h: $P < 2.2e-16$; NPAS4 Q4: 0 vs 2 h $P < 2.2e-16$, 2 vs 10 h: $P < 2.2e-16$, 0 vs 10 h: $P < 2.2e-16$. e. Boxplots of sBLISS-seq DeSeq2 normalized counts in 0 h, 2 h, 10 h ‘less active’, and 10 h ‘still active’ samples at activity-inducible regulatory elements, subset by quartiles of NPAS4

binding. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). Each boxplot represents an individual replicate consisting of an individual mouse. Boxplots of average signal across all replicates are shown in Fig. 4b. f. Aggregate plots (bin size 50 bp) of average SAR-seq[32] coverage ± s.e.m. (n = 3) at CTCF binding sites, all NPAS4 binding sites, and NPAS4 vs FOS binding sites. See Extended Data Fig. 8i for definition of NPAS4 vs FOS binding sites.

Source Data

Extended Data Fig. 10 NPAS4-NuA4 co-binds the genome with DNA repair sensors MRE11 and RAD50 in stimulated neurons.

a. Aggregate plot of average MRE11 CUT&RUN coverage (fragment depth per bp per peak) ± s.e.m. in KA-stimulated nuclei isolated from *Mre11*$^{KO/fl}$ infected with either ΔCre-GFP virus (Control) or Cre-mCherry (*Mre11* cKO) at MRE11 binding sites (17,084 sites). MRE11 binding sites were determined by the SEACR peak calling algorithm. IgG indicates the CUT&RUN signal from a nonspecific IgG control and represents the average IgG across both ΔCre-GFP and Cre conditions. MRE11 (n = 3 Cre and n = 3 ΔCre-GFP), IgG (n = 3 Cre and n = 3 ΔCre-GFP). 1 mouse per replicate. ***P < 2.2e-16. P values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests. b. Integrative Genomics Viewer browser image displaying CUT&RUN signal for NPAS4, MRE11 (Cre or ΔCre), and IgG (Cre or ΔCre) in 2 h KA-stimulated nuclei at the *Bdnf* gene. c. Aggregate CUT&RUN signal for NPAS4, RAD50 and MRE11 in 0 h vs 2 h stimulated hippocampal tissue at NPAS4-binding sites. Each NPAS4-binding site is represented as a single horizontal line centered at the peak summit and extended out ± 1 kb. Intensity of color correlates with sequencing signal as indicated by the scale bar for each factor (0 to 50 read-depth normalization). MRE11(n = 2), RAD50 (n = 4), NPAS4 (n = 5). 3-5 mice pooled per replicate. d. Venn diagram of overlaps between binding sites of MRE11, RAD50 and NPAS4 in 2 h KA-stimulated neurons. Peaks for each factor were determined using the SEACR peak calling algorithm and represent peaks found reproducibly across replicates (2 of 2 MRE11 replicates, 3 of 4 RAD50 replicates, and 4 of 5 NPAS4 replicates). 3-5 mice pooled per replicate. e. Most significant motifs enriched in MRE11

CUT&RUN peaks (2 h KA-stimulated). Motif enrichment was performed on 1 kb peaks extended 500 bp up and downstream from the peak maxima. Notable motifs include the NPAS4/bHLH-PAS motif, the AP1 family, and CTCF motifs. f. Aggregate of average CUT&RUN coverage (fragment depth per bp per peak) ± s.e.m. at all regulatory elements, subset by quartiles of NPAS4 CUT&RUN signal.Q1 = 44,864 sites, Q4 = 7,378 sites. MRE11(0 h and 2 h, n = 2), RAD50 (0 h, n = 3), RAD50 (2 h, n = 4). ***$P < 2.2e\text{-}16$; **$P = 5.78e\text{-}06$. P values by unpaired, two-tailed Wilcoxon rank-sum tests. 3-5 mice pooled per replicate.

Extended Data Fig. 11 NPAS4-NuA4 recruits repair factors to chromatin in stimulated neurons.

a. Boxplots of average IgG-normalized MRE11 and RAD50 CUT&RUN signal (left) or ATAC-seq normalized counts (right) at activity-inducible elements plotted, subset by quartiles of NPAS4 CUT&RUN signal. The sites displayed were selected to have equivalent or higher inducible ATAC-seq signal in low quartiles (Q1) compared to top NPAS4-binding sites (Q4). Q1 = 111 sites, Q4 = 200 sites. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). MRE11(0 h and 2 h, n = 2), RAD50 (0 h, n = 3), RAD50 (2 h, n = 4), ATAC-seq (0 and 2 h, n = 3). 3-5 mice pooled per replicate. MRE11 2 h Q1 vs Q4: ***$P < 2.2e\text{-}16$; RAD50 2 h Q1 vs Q4: ***$P < 2.2e\text{-}16$; ATAC 2 h NPAS4 Q1 vs NPAS4 Q4: $P = 0.105$. P values by unpaired, two-tailed Wilcoxon rank-sum tests. b. Boxplots of average IgG-normalized MRE11 and RAD50 CUT&RUN signal (left) or H3K27ac normalized counts (right) at activity-inducible elements plotted, subset by quartiles of NPAS4 CUT&RUN signal. The sites displayed were selected to have equivalent or higher inducible H3K27ac signal in low quartiles (Q1) compared to top NPAS4-binding sites (Q4). Q1 = 79 sites, Q4 = 88 sites. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). MRE11(0 h and 2 h, n = 2), RAD50 (0 h, n = 3), RAD50 (2 h, n = 4), ATAC-seq (0 and 2 h, n = 3). 3-5 mice pooled per replicate. MRE11 2 h Q1 vs Q4: ***$P < 2.2e\text{-}16$; RAD50 2 h Q1 vs Q4: **$P = 2.603e\text{-}12$; H3K27ac 2 h Q1 vs Q4: $P = 0.0567$. P values by unpaired, two-tailed Wilcoxon rank-sum tests. c. Experimental design used to isolate Cre-mCherry- and ΔCre-GFP-infected nuclei using florescence-activated cell

sorting (FACS). Nuclei are subsequently used for CUT&RUN of EP400, MRE11. d. FACS of Cre-mCherry-positive nuclei (gated from DRAQ5, a fluorescent DNA dye used to identify nuclei). Left panel shows the histogram distribution of Cre-mCherry signal in infected tissue relative to an uninfected control tissue sample. Middle panel shows the sorting scheme, with negative events defined by a DRAQ5-stained, uninfected tissue sample run in parallel on the day of sorting. Right panel shows the FITC vs mCherry signal for all DRAQ5+ nuclei and demonstrates no doubly-infected cells in the samples. See Supplementary Fig. 2 for gating scheme. e. FACS of ΔCre-GFP nuclei (gated from DRAQ5+ nuclei). Left panel shows the histogram distribution of ΔCre-GFP signal in infected tissue relative to an uninfected control tissue sample. Middle panel shows the sorting scheme, with negative events defined by a DRAQ5-stained, uninfected tissue sample run in parallel on the day of sorting. Right panel shows the FITC vs mCherry signal for all DRAQ5+ nuclei and demonstrates no doubly-infected cells in the samples. See Supplementary Fig. 2 for gating scheme. f. Boxplots of average MRE11 CUT&RUN normalized counts in Cre or ΔCre-infected hippocampi of *Npas4*$^{fl/fl}$ mice at activity-inducible sites, subset by quartiles of NPAS4 binding. Q1 = 44,864 sites, Q4 = 7,378 sites. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). MRE11 CRE (n = 3), MRE11 ΔCre (n = 3). 2-3 mice pooled per replicate. MRE11 Q1: $P = 1$, Q4: ***$P < 2.2e\text{-}16$. *P* values by unpaired, one-tailed Wilcoxon rank-sum tests (ΔCre > Cre). g. Boxplots of average EP400 CUT&RUN normalized counts in Cre or ΔCre-infected hippocampi of *Npas4*$^{fl/fl}$ mice at activity-inducible sites, subset by quartiles of NPAS4 binding. Q1 = 44,864 sites, Q4 = 7,378 sites. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). EP400 ΔCre (n = 3), EP400 Cre (n = 4). 2-3 mice pooled per replicate. EP400 Q1: $P = 1$, Q4: ***$P < 2.2e\text{-}16$. *P* values by unpaired, one-tailed Wilcoxon rank-sum tests (ΔCre > Cre). h. Integrative Genomics Viewer browser image displaying CUT&RUN signal for MRE11 and EP400 in *Npas4*-cKO (Cre) or Control (ΔCre) in 2 h KA-stimulated nuclei at the *Rgs7bp* gene. MRE11 CRE (n = 3), MRE11 ΔCre (n = 3), EP400 ΔCre (n = 3), EP400 Cre (n = 4). 2-3 mice pooled per replicate.

Extended Data Fig. 12 Loss of NPAS4-NuA4 increases DNA breaks across the neuronal genome.

a. Boxplots of sBLISS-seq normalized counts in nuclei isolated from *Npas4*$^{fl/fl}$ mice injected with Cre-mCherry (*Npas4*-cKO) or ΔCre-GFP virus (Control) at activity-inducible regulatory elements, subset by quartiles of NPAS4 CUT&RUN signal (see Methods, Extended Data Fig. 7a). Q1 = 1,017 sites, Q4 = 764 sites. Data plotted includes all datapoints coming from 5 replicates per genotype; no averaging across replicates was performed. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). ***$P < 2.2e-16$ P values by unpaired, one-tailed Wilcoxon rank-sum tests (Cre > ΔCre). b. Boxplots of sBLISS-seq normalized counts in nuclei isolated from wild-type mice injected with Cre-mCherry or ΔCre-GFP virus at activity-inducible regulatory elements, subset by quartiles of NPAS4 CUT&RUN signal. Q1 = 1,017 sites, Q4 = 764 sites. Data plotted includes all datapoints coming from 3 replicates per genotype; no averaging across replicates was performed. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). Q1: $P = 1$; Q4 $P = 0.995$. P values by unpaired, one-tailed Wilcoxon rank-sum tests (Cre > ΔCre). c. Boxplots of sBLISS-seq normalized counts in nuclei isolated from *Npas4*$^{fl/fl}$ mice injected with Cre-mCherry (*Npas4*-cKO) or ΔCre-GFP virus (Control) at all regulatory elements, subset by quartiles of NPAS4 CUT&RUN signal (see Methods for regulatory site definition). Q1 = 44,864 sites, Q4 = 7,378 sites. Data plotted includes all datapoints coming from 5 replicates per genotype. Boxes represent the interquartile range with line at the median. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). ***$P<2.2e-16$. P values by unpaired, one-tailed Wilcoxon rank-sum tests (Cre > ΔCre). d. Boxplots of sBLISS-seq DeSeq2 normalized counts in nuclei isolated from wild-type injected with Cre-mCherry or ΔCre-GFP virus at all regulatory elements, subset by quartiles of NPAS4 CUT&RUN signal (see Methods for regulatory site definition). Data plotted includes all datapoints coming from 3 replicates per genotype. Boxplot shows median (line), IQR (box), 1.5x IQR (whiskers), notches indicate median ± 1.58× IQR/sqrt(n). Q1: $P = 1$; Q4 $P = 0.1092$. P values by unpaired, one-tailed Wilcoxon rank-sum tests (Cre > ΔCre). e. Average ± s.e.m. of genome-wide breaks in hippocampal nuclei

isolated from *Tip60*$^{fl/fl}$ mice (0 and 2 h; n = 3, 10 h; n = 4) infected with Cre or ΔCre virus. To compare across samples, input reads were downsampled to the lowest value among all conditions, and numbers of unique DNA breaks are shown. *Tip60*$^{fl/fl}$: 0 h: $P = 0.0017$, 2 h: $P = 0.013$, 10 h: $P = 0.158$. *P* values by two-tailed, unpaired t-tests. f. Cleaved caspase 3 (apoptosis marker) staining in *Npas4*$^{fl/fl}$ mice injected with AAV to express Cre-mCherry. Representative image from 3 animals. Scale bar = 300 μm. g. Cleaved caspase 3 staining in *Tip60*$^{fl/fl}$ mice injected with AAV to express Cre-mCherry. Representative image from 3 animals. Scale bar = 300 μm.

Source Data

Extended Data Fig. 13 Mutational Analysis at NPAS4-NuA4 sites during aging and *Npas4 KO* lifespan analysis.

a. FACS sorting scheme to isolate NeuN-expressing neurons from aging *Npas4* wild-type vs KO hippocampal tissue. Sorted cells from NeuN+ and NeuN- gates were re-analyzed for purity. See Supplementary Fig. 2 for gating scheme. b. Average normalized gene expression ± s.e.m. (relative to housekeeping gene *Gapdh*) of marker genes for neurons (NeuN+) or glial cell types (NeuN-). Each dot represents data from an individual mouse collected across 2 independent sorting days. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. *P* values by unpaired, two-tailed t-tests with Benjamini-Hochberg multiple hypothesis correction. Individual *P* values and replicates as follows: *Grin1* (n = 6): $P = 2.06\text{e-}05$, *Grin2b* (n = 3): $P = 5.97\text{e-}03$, *Syapsin1* (n = 3): 5.72e-04, *Rbfox3* (n = 6): 3.89e-04, *Npas4* (n = 3): $P = 5.21\text{e-}03$, *S100b* (n = 6): $P = 3.85\text{e-}06$, *Gfap* (n = 3): $P = 1.89\text{e-}02$, *Aldh1l1* (n = 3): $P = 9.50\text{e-}03$, *Mog* (n = 3): $P = 2.29\text{e-}04$, *Pdgrfa* (n = 6): $P = 1.16\text{e-}05$. c. Diagram of the SiMSen-seq amplicon sequencing approach used to identify somatic mutations that occur in individual DNA templates. d. Diagram of positive control experiment (top) used to test SiMSen-seq mutation detection method. Cultured cortical neurons were infected with lentivirus to express Cas9 with or without guide RNAs (*Bdnf, Fos, Inhba, Scg2*). SiMSen-seq was used to detect mutations at the gRNA cut sites at *Inhba* and *Bdnf*. Mutation frequency (that is, the number of insertions/deletion per UMI family) is plotted across a 40 bp window surrounding the cut site. e. Violin plots of read depth-normalized sequencing signal for NPAS4,

MRE11, RAD50, H3K27ac CUT&RUN, γH2AX ChIP-seq and ATAC-seq in 2 h stimulated neurons at NPAS4-bound (24) and No NPAS4 (11) sites selected for mutational analysis. Each point represents a site. Inducible H3K27ac and Inducible ATAC plots show the distribution of fold changes (2 vs 0 h) in ATAC-seq and H3K27ac at these same sites. Dashed line represents the median; solid lines represent quartiles. ***$P < 0.0005$; **$P < 0.005$. P values by unpaired, two-tailed Wilcoxon rank sum tests. NPAS4: $P = 0.002$, MRE11: $P = 0.0009$, RAD50: $P = 1.33\text{e-}04$, γH2AX $P = 0.0027$, Percent GC $P = 0.76$; Percent AT $P = 0.76$; Inducible H3K27ac $P = 0.25$, H3K27ac $P = 0.07$; ATAC $P = 0.19$; Inducible ATAC $P = 0.53$. Replicates per factor included in Supplementary Table 2. f. Genomic annotations of NPAS4-bound (24) vs No NPAS4 (11) sites. g. Left panel: Average SNV frequency ± s.e.m. in NPAS4-Bound and No NPAS4 in young (3-month-old) animals. Each point represents a single site sampled from a mouse and data from 4 mice are shown. *$P < 0.05$; **$P < 0.005$, ***$P < 0.001$; T>A $P = 0.99$, T>G $P = 0.99$, T>C $P = 0.0049$, C>A $P = 0.027$, C>G $P = 0.99$, C>T $P = 1.85\text{e-}09$. P values by a one-way ANOVA with Holm-Sidak's correction for multiple hypothesis testing. Right panel: Average Insertion/Deletion frequency ± s.e.m. in NPAS4-Bound and No NPAS4 sites in young (3-month-old) animals. Each point represents a single site sampled from a mouse and data from 4 mice are shown. ***$P = 2.96\text{e-}05$. P value by an unpaired, two-tailed Wilcoxon rank-sum test. h. Lifespan analysis on *Npas4* wild-type (n = 25) vs *Npas4* KO (n = 27) male littermates. Median lifespan KO = 12 months; Median lifespan of wild-type not determined. $P = 8.67\text{e-}06$ by two-tailed Gehan-Breslow-Wilcoxon test; $P = 1.37\text{e-}06$ by two-tailed log-rank Mantel-Cox test. i. Lifespan analysis on *Npas4* wild-type (n = 28) vs *Npas4* KO (n = 37) female littermates. Median lifespan KO = 11 months; Median lifespan of wild-type not determined. $P = 1.01\text{e-}06$ by two-tailed Gehan-Breslow-Wilcoxon test; $P = 8.48\text{e-}08$ by two-tailed log-rank Mantel-Cox test. j. Lifespan analysis on *Npas4* wild-type (n = 16, $Npas4^{+/+}$; *Camk2a-Cre+*; $Sun1^{fl/+}$) vs *Npas4* cKO (n = 9, $Npas4^{fl/fl}$; *Camk2a-Cre+*; $Sun1^{fl/+}$). Median lifespan cKO = 21.46 months; Median lifespan of wild-type = 29 months. $P = 0.049$ by two-tailed Gehan-Breslow-Wilcoxon test; $P = 0.19$ by a two-tailed log-rank Mantel-Cox test.

Source Data

Supplementary information

Supplementary Figures

Supplementary Fig. 1: full scan images for western blots in Fig. 1b and Extended Data Figs. 1a,d–f,h–j, 2a,b,d and 4d. Supplementary Fig. 2: gating strategy for Fig. 5d and Extended Data Figs. 11d,e and 13a.

Reporting Summary

Peer Review File

Supplementary Table 1

Summary of peptides obtained in mass spectrometry experiments. For raw data, please see PRIDE repository under accession number PXD038718. Benjamini–Hochberg adjusted *P* values reported were calculated using EdgeR (v.3.28.1).

Supplementary Table 2

Table of all sequencing data with replicate information.

Supplementary Table 3

Hippocampal snRNA-seq gene targets of NPAS4 and TIP60 across cell types. Bonferroni-adjusted *P* values were calculated using unpaired, two-tailed Wilcoxon rank-sum tests implemented in Seurat (v.3).

Supplementary Table 4

Genomic locations for hippocampal regulatory elements and inducible ATAC-seq and H3K27ac CUT&RUN. Benjamini–Hochberg-adjusted *P* values reported were calculated using DeSeq2 (v.1.26.0).

Supplementary Table 5

Regions and primers used in mutation analysis.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 4

Source Data Fig. 5

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 12

Source Data Extended Data Fig. 13

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article′s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article′s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Pollina, E.A., Gilliam, D.T., Landau, A.T. *et al.* A NPAS4–NuA4 complex couples synaptic activity to DNA repair. *Nature* **614**, 732–741 (2023). https://doi.org/10.1038/s41586-023-05711-7

- Received: 28 May 2021
- Accepted: 05 January 2023
- Published: 15 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05711-7

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

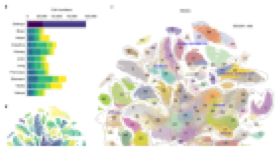
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

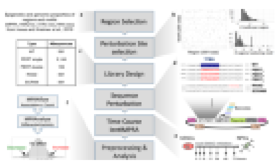
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05711-7



Systematic identification of cell-fate regulatory programs using a single-cell atlas of mouse development

11 July 2022

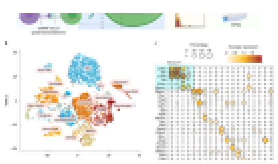
Lijiang Fei, Haide Chen, … Guoji Guo



Massively parallel reporter perturbation assays uncover temporal regulatory architecture during neural differentiation

21 March 2022

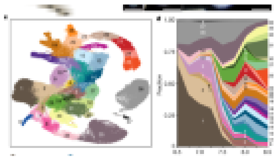
Anat Kreimer, Tal Ashuach, … Nadav Ahituv



CellComm infers cellular crosstalk that drives haematopoietic stem and progenitor cell development

12 April 2022

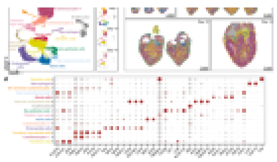
Edroaldo Lummertz da Rocha, Caroline Kubaczka, … George Q. Daley



[A single-cell molecular map of mouse gastrulation and early organogenesis]

20 February 2019

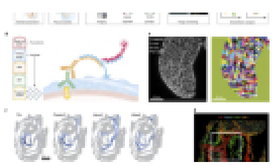
Blanca Pijuan-Sala, Jonathan A. Griffiths, … Berthold Göttgens



[Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis]

19 March 2021

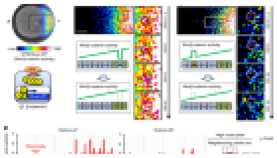
Madhav Mantri, Gaetano J. Scuderi, … Iwijn De Vlaminck



[Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis]

06 September 2021

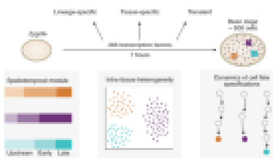
T. Lohoff, S. Ghazanfar, … J. C. Marioni



[Cell competition corrects noisy Wnt morphogen gradients to achieve robust patterning in the zebrafish embryo](#)

17 October 2019

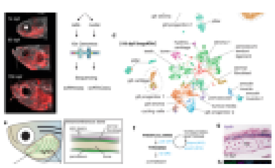
Yuki Akieda, Shohei Ogamino, … Tohru Ishitani



[The long and winding road of development: a coordinated song of transcription factors](#)

26 July 2021

Aly Makhlouf & Marta N. Shahbazi



[Lifelong single-cell profiling of cranial neural crest diversification in zebrafish](#)

10 January 2022

Peter Fabian, Kuo-Chang Tseng, … J. Gage Crump

- Article
- Open Access
- Published: 08 February 2023

Dissecting cell identity via network inference and in silico gene perturbation

- Kenji Kamimoto,
- Blerta Stringa,
- Christy M. Hoffmann,
- Kunal Jindal,
- Lilianna Solnica-Krezel &
- …
- Samantha A. Morris

Show authors

Nature **volume 614**, pages 742–751 (2023)

- 27k Accesses
- 319 Altmetric
- Metrics details

Abstract

Cell identity is governed by the complex regulation of gene expression, represented as gene-regulatory networks[1]. Here we use gene-regulatory networks inferred from single-cell multi-omics data to perform in silico transcription factor perturbations, simulating the consequent changes in cell identity using only unperturbed wild-type data. We apply this machine-learning-based approach, CellOracle, to well-established paradigms—mouse and human haematopoiesis, and zebrafish embryogenesis—and we correctly model reported changes in phenotype that occur as a result of transcription factor perturbation. Through systematic in silico transcription factor perturbation in the developing zebrafish, we simulate and experimentally

validate a previously unreported phenotype that results from the loss of *noto*, an established notochord regulator. Furthermore, we identify an axial mesoderm regulator, *lhx1a*. Together, these results show that CellOracle can be used to analyse the regulation of cell identity by transcription factors, and can provide mechanistic insights into development and differentiation.

Main

The expansion of single-cell technologies into perturbational omics is enabling the development of methods to characterize cell identity. For example, single-cell RNA sequencing (scRNA-seq) coupled with pooled CRISPR screens offers much promise for analysing the genetic regulation of cell identity[2,3,4], but cannot be readily used in many biological contexts. Computational methods to simulate single-cell phenotypes after perturbation are emerging, although many approaches still require experimental perturbation data for model training, and thus their scale and application are limited[5]. Moreover, previous deep-learning-based models represent a 'black box', which restricts the interpretation of gene-regulatory mechanisms that underlie the simulated biological events. In this respect, gene-regulatory network (GRN) modelling approaches are promising as they reconstruct systematic gene–gene associations from unperturbed single-cell omics data[6,7,8,9,10,11]. However, previous methods for analysing GRNs largely focus on the static network structure, and determining how a static GRN governs cell identity during dynamic biological processes therefore remains a challenge. Scalable and interpretable approaches are required to understand how gene-regulatory mechanisms relate to observed complex single-cell phenotypes.

Here we present a strategy that overcomes these limitations by combining computational perturbation with GRN modelling. CellOracle integrates multimodal data to build custom GRN models that are specifically designed to simulate shifts in cell identity following transcription factor (TF) perturbation, providing a systematic and intuitive interpretation of context-dependent TF function in regulating cell identity. We apply CellOracle to well-characterized biological systems: haematopoiesis in mice and humans; and the differentiation of axial mesoderm into notochord and prechordal plate in zebrafish. In haematopoiesis, we show that CellOracle recapitulates

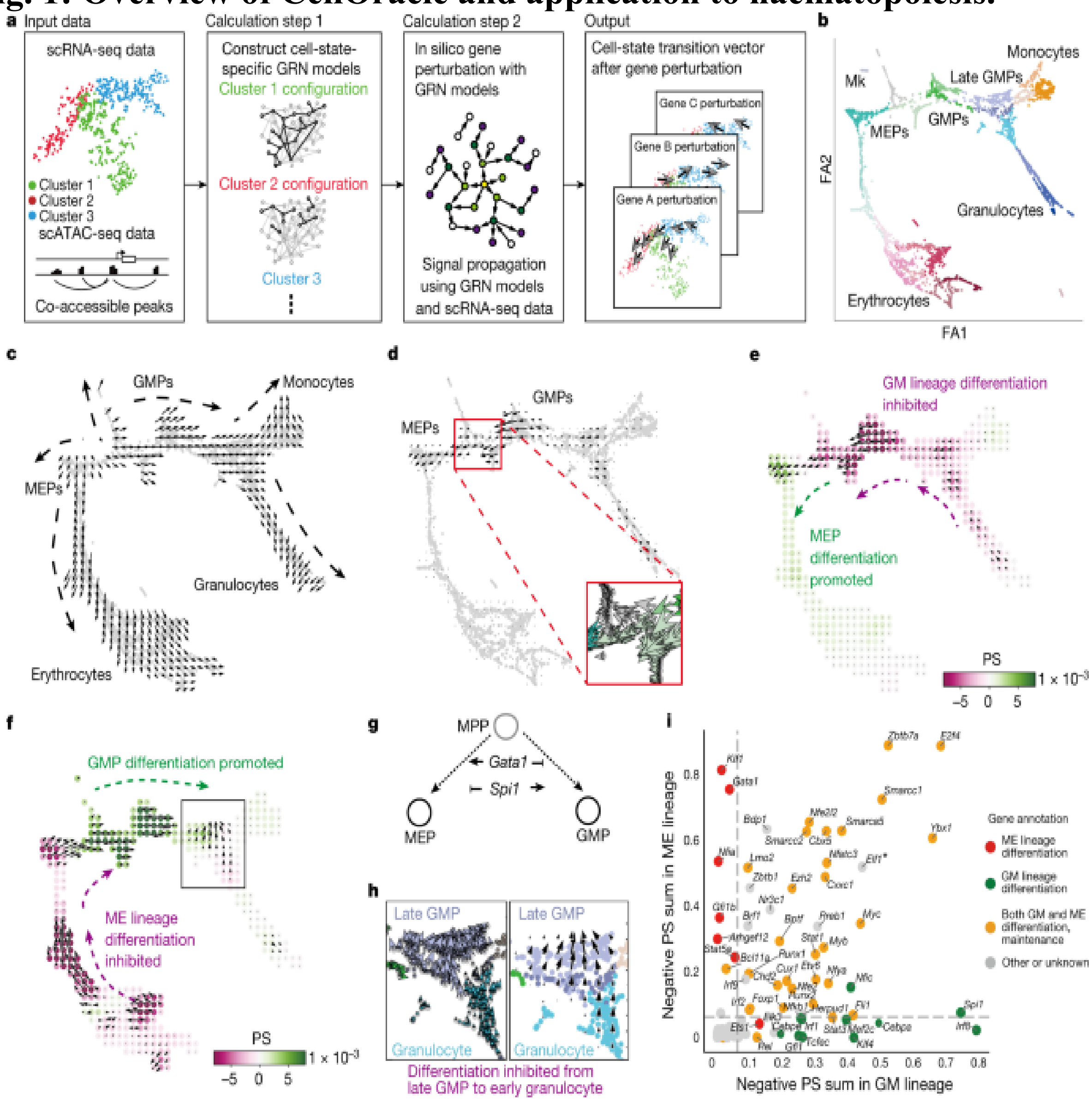
well-known cell fate regulation governed by TFs. Furthermore, we apply CellOracle to systematically perturb TFs across zebrafish development, recovering known and putative regulators of cell identity. Focusing on axial mesoderm, we predict and validate a prechordal plate phenotype after loss of function (LOF) of the prototypical notochord regulator, *noto*. Moreover, we also simulate and validate a role for the TF *lhx1a* in the development of axial mesoderm. Together, these results show that CellOracle can be used to infer and interpret cell-type-specific GRN configurations at high resolution, enabling mechanistic insights into the regulation of cell identity. CellOracle code and documentation are available at https://github.com/morris-lab/CellOracle and data can be explored at https://celloracle.org.

In silico gene perturbation using CellOracle

To gain mechanistic insight into the regulation of cell identity, we developed an in silico strategy to simulate changes in cell identity upon TF perturbation. CellOracle uses custom GRN modelling (Extended Data Fig. 1a) to simulate global downstream shifts in gene expression following knockout (KO) or overexpression of TFs. These simulated values are converted into a vector map of transitions in cell identity, which enables simulated changes in cell identity to be intuitively visualized within a low-dimension space (Fig. 1a and Methods). In silico perturbation involves four steps. (1) Cell-type- or cell-state-specific GRN configurations are constructed using cluster-wise regularized linear regression models with multi-omics data. (2) Using these GRN models, shifts in target gene expression in response to TF perturbation are calculated. This step applies the GRN model as a function to propagate the shift in gene expression rather than the absolute gene expression value, representing the signal flow from TF to target gene. This signal is propagated iteratively to calculate the broad, downstream effects of TF perturbation, allowing the global transcriptional 'shift' to be estimated (Extended Data Fig. 1b–d). (3) The cell-identity transition probability is estimated by comparing this shift in gene expression to the gene expression of local neighbours. (4) The transition probability is converted into a weighted local average vector to represent the simulated directionality of cell-state transition for each cell following perturbation of candidate TFs. In the final calculation step, the multi-dimensional gene expression shift vector is reduced to a two-dimensional (2D) vector,

allowing for more robust predictions against noise (Extended Data Fig. 1e). We purposefully limit the simulation output data to a 2D vector representing the predicted shift in cell identity because our goal is to model changes in identity rather than predicting absolute changes in gene expression levels. Further details of the CellOracle algorithm are provided in the Methods, including validation of the range of simulated values; null or randomized model analysis; and hyperparameter evaluation (Supplementary Figs. 2–10).

Fig. 1: Overview of CellOracle and application to haematopoiesis.



a, Simulation of cell-state transitions in response to TF perturbation. First, CellOracle constructs custom transcriptional GRNs using scRNA-seq and

scATAC-seq data (left). Accessible promoter and enhancer peaks from scATAC-seq data are then combined with scRNA-seq data to generate cluster-specific GRN models (middle). CellOracle simulates the change in cell state in response to a TF perturbation, projecting the results onto the cell trajectory map (right). **b**, Force-directed graph of 2,730 myeloid progenitor cells from Paul et al.[16]. Twenty-four cell clusters (Louvain clustering) were organized into six main cell types. Mk, megakaryocytes. **c**, Differentiation vectors for each cell projected onto the force-directed graph. **d**, CellOracle simulation of cell-state transition in *Spi1* KO simulation. Summarized cell-state transition vectors projected onto the force-directed graph. Vectors for each cell are shown in the inset. **e**, *Spi1* KO simulation vector field with perturbation scores (PSs). **f**, *Gata1* KO simulation with perturbation scores. **g**, Schematic of *Spi1–Gata1* lineage switching. MPP, multipotent progenitor. **h**, Detail of *Gata1* simulation for the granulocyte branch. Left, cell-state transition vectors for each cell. Right, summarized vectors. **i**, Systematic KO simulation result of 90 TFs in the GM and ME lineage is summarized as a scatter plot of the sum of negative perturbation scores (shown in log scale). Dashed lines represent cut-off values corresponding to false-positive rate (FPR) = 0.01. Genes are classified into four categories on the basis of their previously reported functions (Supplementary Table 2). The asterisk refers to Supplementary Fig. 11, where we expand on the predicted phenotype. All scores can be explored through our web application (https://celloracle.org).

GRN inference and benchmarking with CellOracle

The CellOracle GRN model must represent regulatory connections as a directed network edge to support signal propagation in response to TF perturbation. Thus, we developed a custom GRN modelling method motivated by previous approaches that incorporate promoter and TF-binding information with scRNA-seq data to infer a directional GRN[7] (Extended Data Fig. 1a and Methods). First, using single-cell chromatin accessibility data (single-cell assay for transposase-accessible chromatin using sequencing; scATAC-seq), we incorporate flexible promoter and enhancer regions, encompassing proximal and distal regulatory elements. This initial step uses the transcriptional start site (TSS) database (http://homer.ucsd.edu/) and Cicero, an algorithm that identifies co-accessible scATAC-seq peaks, to distinguish accessible promoters and

enhancers[12]. The DNA sequence of these elements is then scanned for TF-binding motifs, generating a 'base GRN structure' of all potential regulatory interactions in the species of interest (Extended Data Fig. 1a, left). This process is beneficial as it narrows the scope of possible regulatory candidate genes before model fitting (below) and helps define the directionality of regulatory edges in the GRN. To support GRN inference without requiring sample-specific scATAC-seq datasets, we have assembled a base GRN from a mouse scATAC-seq atlas[13]. We have also created general promoter base GRNs for ten commonly studied species (Supplementary Table 1 and Methods). These base GRNs are built into the CellOracle library and provide an alternative solution when scATAC-seq data are unavailable.

In the second step of CellOracle GRN inference, we use scRNA-seq data to identify active connections in the base GRN, generating cell-type- or cell-state-specific GRN configurations for each cluster. In this step, we build a machine-learning model to predict the expression of target genes on the basis of TF expression (Extended Data Fig. 1a, right). Because CellOracle uses genomic sequences and information on TF-binding motifs to infer the base GRN structure and directionality, it does not need to infer the causality or directionality of the GRN from expression data. This approach allows CellOracle to adopt a relatively simple modelling method for GRN inference —a regularized linear machine-learning model. Crucially, this strategy enables the above signal propagation to simulate TF perturbation. To support the use of a linear model, the gene expression matrix of scRNA-seq data is divided into several clusters in advance so that a single data unit for each fitting process represents a linear relationship rather than non-linear or mixed regulatory relationships. Furthermore, a Bayesian or bagging strategy enables the certainty of connection to be presented as a distribution; this allows weak or insignificant connections to be removed from the base GRN (Extended Data Fig. 1a, right), producing a cell-type- or cell-state-specific GRN configuration.

To benchmark our GRN inference method, we generated a comprehensive transcriptional ground-truth GRN using 1,298 chromatin immunoprecipitation followed by sequencing (ChIP–seq) datasets for 80 regulatory factors across 5 different tissues[14]. In addition to benchmarking against diverse GRN inference algorithms, we also assessed the performance

of our approach using different base GRNs, data sources and cell downsampling (Extended Data Fig. 2). Inference performance as assessed by the area under the receiver operating characteristic (AUROC) ranged from 0.66 to 0.85 for the promoter base GRN and 0.73 to 0.91 for the scATAC-seq base GRN. Altogether, this benchmarking demonstrates the accuracy of our transcriptional GRN modelling method with a diverse range of data sources. Combined with our signal propagation strategy, CellOracle can effectively interrogate network biology and cell-identity dynamics through in silico perturbation.

GRN analysis and TF KO in haematopoiesis

For validation, we aimed to reproduce known TF regulation of mouse haematopoiesis, a well-characterized differentiation paradigm[15], by applying CellOracle to a 2,730-cell scRNA-seq atlas of myeloid progenitor differentiation[16] (Fig. 1b and Extended Data Fig. 3a). We constructed GRN models for each of the 24 myeloid clusters identified, representing megakaryocyte and erythroid progenitors (MEPs) and granulocyte–monocyte progenitors (GMPs), differentiating toward erythrocytes, megakaryocytes, monocytes and granulocytes (Fig. 1c). To test whether the CellOracle simulation could recapitulate known TF regulation of cell identity, we performed in silico gene perturbation using the inferred GRNs, and compared the CellOracle KO simulation results with previous biological knowledge and ground-truth KO data.

First, *Spi1* (also known as *PU.1*) and *Gata1* KO simulation is used to illustrate the CellOracle in silico perturbation analysis. The TF perturbation simulation is visualized as a vector map on the 2D trajectory space (Fig. 1d and Supplementary Video 1), representing a potential shift in cell identity in response to TF perturbation. To enable the simulation results to be assessed systematically and objectively, we also devised a ‘perturbation score’ metric, which compares the directionality of the perturbation vector to the natural differentiation vector (Extended Data Fig. 4). A negative perturbation score suggests that TF KO delays or blocks differentiation (Extended Data Fig. 4b–d, purple). Conversely, a positive perturbation score suggests that the differentiation and KO simulation vectors share the same direction, indicating that loss of TF function promotes differentiation (Extended Data

Fig. 4b–d, green). *Spi1* KO simulation yielded positive perturbation scores for MEPs, whereas GMPs had negative perturbation scores (Fig. 1e), suggesting that *Spi1* KO inhibits GMP differentiation and promotes MEP differentiation. Inverse perturbation score distributions were produced for the *Gata1* KO simulation (Fig. 1f). Comparing these predictions to previous reports[17,18]: PU.1 directs commitment to the neutrophil and monocyte lineages[19,20], whereas GATA1 promotes the differentiation of erythroid cells[21] and eosinophil granulocytes[22,23,24]. Overall, CellOracle accurately simulated the myeloid lineage switching governed by *Gata1* and *Spi1* (refs. [15,25,26,27]; Fig. 1g), including a relatively mild *Gata1* KO phenotype in early granulocyte differentiation (Fig. 1h), which cannot be inferred from the low levels of *Gata1* expression in granulocytes (Extended Data Fig. 3d). However, CellOracle did not detect a previously reported depletion of erythrocyte progenitors after *Spi1* KO[27,28], probably owing to changes in cell proliferation that are not predicted by the method.

We next evaluated eight additional TFs that have established roles in myeloid differentiation: *Klf1* (also known as *Eklf)*, *Gfi1b*, *Fli1*, *Gfi1*, *Gata2*, *Lmo2*, *Runx1* and *Irf8* (refs. [15,29]). CellOracle also correctly reproduced their reported KO phenotypes (Extended Data Figs. 5 and 6), which we extended to two additional datasets of mouse and human haematopoiesis (Extended Data Figs. 7 and 8 and Supplementary Figs. 13 and 14). In addition, we scaled up our simulation to all TFs that passed filtering (Methods) to systematically perturb 90 TFs in the dataset in the context of granulocyte–monocyte (GM) and megakaryocyte–erythroid (ME) differentiation. The reported cell-fate-regulatory functions of these TFs fall into three major categories: (1) ME lineage differentiation; (2) GM lineage differentiation; and (3) ME and GM lineage differentiation and maintenance of haematopoietic stem cell (HSC) identity (Supplementary Table 2). We ranked the TFs on the basis of the sum of the negative perturbation score in the KO simulation, representing the potential of a TF potential to promote differentiation (Methods and Extended Data Fig. 3f).

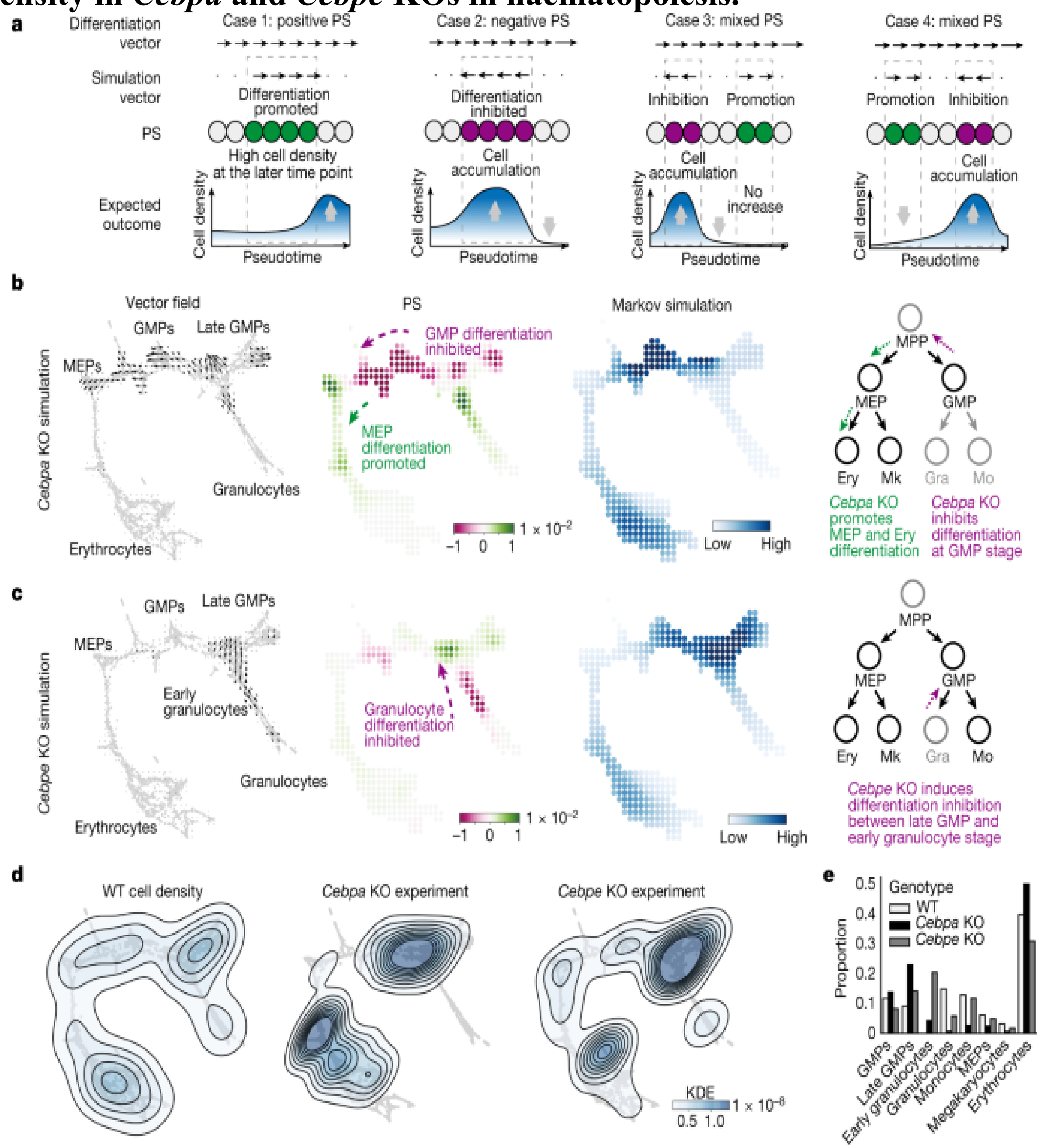
To summarize this systematic TF perturbation, the summed negative perturbation scores are shown on a scatter plot (Fig. 1i). The dashed lines represent cut-off values calculated with a randomized vector (Extended Data Fig. 3g). The distribution of negative perturbation score sums for all TF KOs

was highly consistent with known TF functions in differentiation. For example, TFs involved in ME lineage differentiation are enriched on the top left side of the scatter plot. By contrast, GM differentiation factors are found at the bottom right. TFs that regulate both lineages are located on the top right side, whereas the lower-ranked factors are enriched for TFs that have not been reported to regulate blood differentiation (Fig. 1i). Overall, 85% of the top 30 TFs ranked by this objective, systematic perturbation strategy are reported regulators of myeloid differentiation (Supplementary Table 2). Of the remaining TFs, several have no reported phenotypes in haematopoiesis at present, and therefore represent putative regulators. We note that the negative perturbation score metric does not always convey all information of the vector field, which might oversimplify the role of a TF. For example, *Elf1* has a negative perturbation score in both the ME and the GM lineage, and its function is unclear on the summarized perturbation score plot; however, closer inspection of the vector reproduced its reported phenotype in the ME lineage, highlighting the importance of investigating the simulation output (Supplementary Fig. 11). Finally, we directly compared the output of CellOracle to existing methods for identifying regulatory TFs using gene expression and chromatin accessibility, demonstrating the unique insights into context-dependent TF regulation that CellOracle can provide (Supplementary Figs. 15 and 16).

We further validated CellOracle simulation by focusing on several genes for which experimental KO scRNA-seq data are available: *Cebpa*, *Cebpe* and *Tal1* (refs. [16,30]). *Cebpa* is necessary for the initial differentiation of GMPs, and its loss leads to a marked decrease in differentiated myeloid cells, accompanied by an increase in erythroid progenitors. By contrast, *Cebpe* is not required for initial GMP differentiation, but it is essential for the subsequent maturation of GMPs into granulocytes[16]. Notably, when we compare the simulation results to the experimental KO cell distribution, we must again consider the effects of TF perturbation in the context of natural cell differentiation (Fig. 2a). Thus, we performed a Markov random walk simulation based on the differentiation and simulation vectors to estimate how TF perturbation leads to changes in cell distribution (Supplementary Fig. 17 and Methods). For *Cebpa*, CellOracle simulation predicted that differentiation is inhibited at GMP–late GMP clusters, whereas early erythroid differentiation is promoted (Fig. 2b). The simulation recapitulates

the experimental cell distribution (Fig. 2b,d). For *Cebpe*, CellOracle again correctly modelled the inhibition of differentiation at the entry stage of granulocyte differentiation (Fig. 2c), consistent with experimental KO data (Fig. 2d).

Fig. 2: Validation of CellOracle using experimentally measured cell density in *Cebpa* and *Cebpe* KOs in haematopoiesis.



a, Biological interpretation of perturbation scores (estimation of cell density based on perturbation score). Case 1: the differentiation and perturbation simulation vectors share the same direction, indicating a population shift towards a more differentiated identity. Case 2: the two vectors are opposed, suggesting that differentiation is inhibited. Case 3: predicted inhibition precedes promotion; thus, cells will be likely to accumulate. **b**,**c**, CellOracle *Cebpa* KO (**b**) and *Cebpe* KO (**c**) simulations showing cell-state transition vectors, perturbation scores and estimated cell density (Markov simulation). Right, schematics of simulated phenotype. Ery, erythrocyte. **d**, Ground-truth experimental cell density plot of wild-type (WT) cells, *Cebpa* KO cells and *Cebpe* KO cells in the force-directed graph embedding space. Estimated kernel density data are shown as a contour line on a scatter plot to depict cell density. **e**, Cell-type proportions in the WT and ground-truth KO samples. Gra, granulocyte; KDE, kernel density estimation; Mo, monocyte.

We also analysed a single-cell atlas of mouse organogenesis[30] to simulate the loss of *Tal1* function (Extended Data Fig. 9a–d). CellOracle reproduced the inhibited differentiation of haematoendothelial progenitors in the *Tal1* KO[30] (Extended Data Fig. 9e–h). In addition, CellOracle showed that loss of *Tal1* in later stages of erythroid differentiation does not block cell differentiation (Extended Data Fig. 9i,j), consistent with previous conditional *Tal1* KO experiments at equivalent stages[31]. Together, these results show that CellOracle effectively simulates cell-state-specific TF function, corroborating previous knowledge of the mechanisms that regulate cell fate in haematopoiesis and ground-truth in vivo phenotypes. Furthermore, systematic KO simulations demonstrate that CellOracle enables objective and scalable in silico gene perturbation analysis.

Systematic TF KO simulations in zebrafish

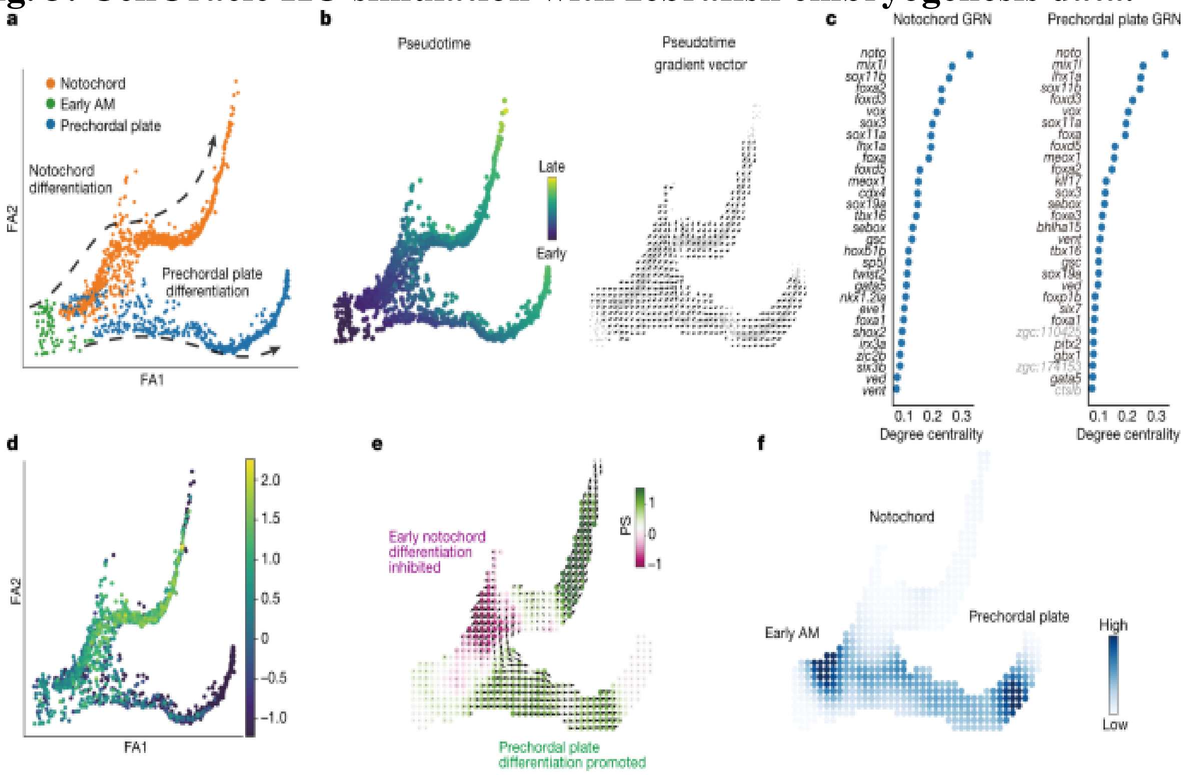
Next, we applied CellOracle to systematically perturb TFs across zebrafish development. We made use of a 38,731-cell atlas of zebrafish embryogenesis published in a study by Farrell et al.[32], comprising 25 developmental trajectories that span zygotic genome activation to early somitogenesis. We first inferred GRN configurations for the 38 cell types and states identified in the Farrell et al. study[32], splitting the main branching trajectory into four sub-branches: ectoderm; axial mesoderm; other

mesendoderm; and germ layer branching point (Extended Data Fig. 10a,b). In the absence of scATAC-seq data, we constructed a base GRN using promoter information from the UCSC database, obtaining information on TF-binding motifs from the *Danio rerio* CisBP motif database (Methods). Our benchmarking has shown that this approach produces reliable GRN inference (Extended Data Fig. 2). After preprocessing and GRN inference, we performed KO simulations for all TFs with inferred connections to at least one other gene ($n = 232$ 'active' TFs; Methods). The results of these simulations across all developmental trajectories can be explored at https://www.celloracle.org.

Our systematic TF KO simulation provides a valuable resource for identifying regulators of early zebrafish development and enables candidates to be prioritized for experimental validation. To further examine this comprehensive perturbation atlas, we focused on axial mesoderm differentiation, spanning 4.3 to 12 h post-fertilization (hpf) (Fig. 3a,b and Extended Data Fig. 10a,b). This midline structure bifurcates into notochord and prechordal plate lineages, representing a crucial patterning axis[33], and has been extensively characterized, in part through large-scale genetic screens[34]. For these lineages, we performed systematic TF KO simulation and network analysis for 232 candidate TFs (Extended Data Fig. 10c). CellOracle ranked *noto*, a well-characterized TF regulator of notochord development, as the top TF on the basis of degree centrality, along with other known regulators of notochord development (Fig. 3c). Degree centrality is a straightforward measure that reports how many edges (genes) are directly connected to a node (TF); highly connected nodes are likely to be essential for a biological process[35,36]. In zebrafish *floating head*$^{n1/n1}$ (*flh*$^{n1/n1}$) mutants, which lack a functional *noto* gene (*noto* is also known as *flh*)[37], axial mesoderm does not differentiate into notochord, and assumes a somitic mesoderm fate instead[38]. *Noto* LOF simulation correctly reproduced the loss of notochord (Fig. 3d–f and Extended Data Fig. 10d–f), in addition to enhanced somite differentiation (Extended Data Fig. 10g–k). Moreover, CellOracle predicted a previously unknown (to our knowledge) consequence of *noto* LOF: enhanced prechordal plate differentiation (Fig. 3e,f). We also noted that later stages of notochord differentiation received a positive perturbation score, indicating that continued expression of *noto* is

not required for notochord differentiation. Alternatively, this finding could suggest that downregulation of *noto* is required for notochord maturation.

Fig. 3: CellOracle KO simulation with zebrafish embryogenesis data.

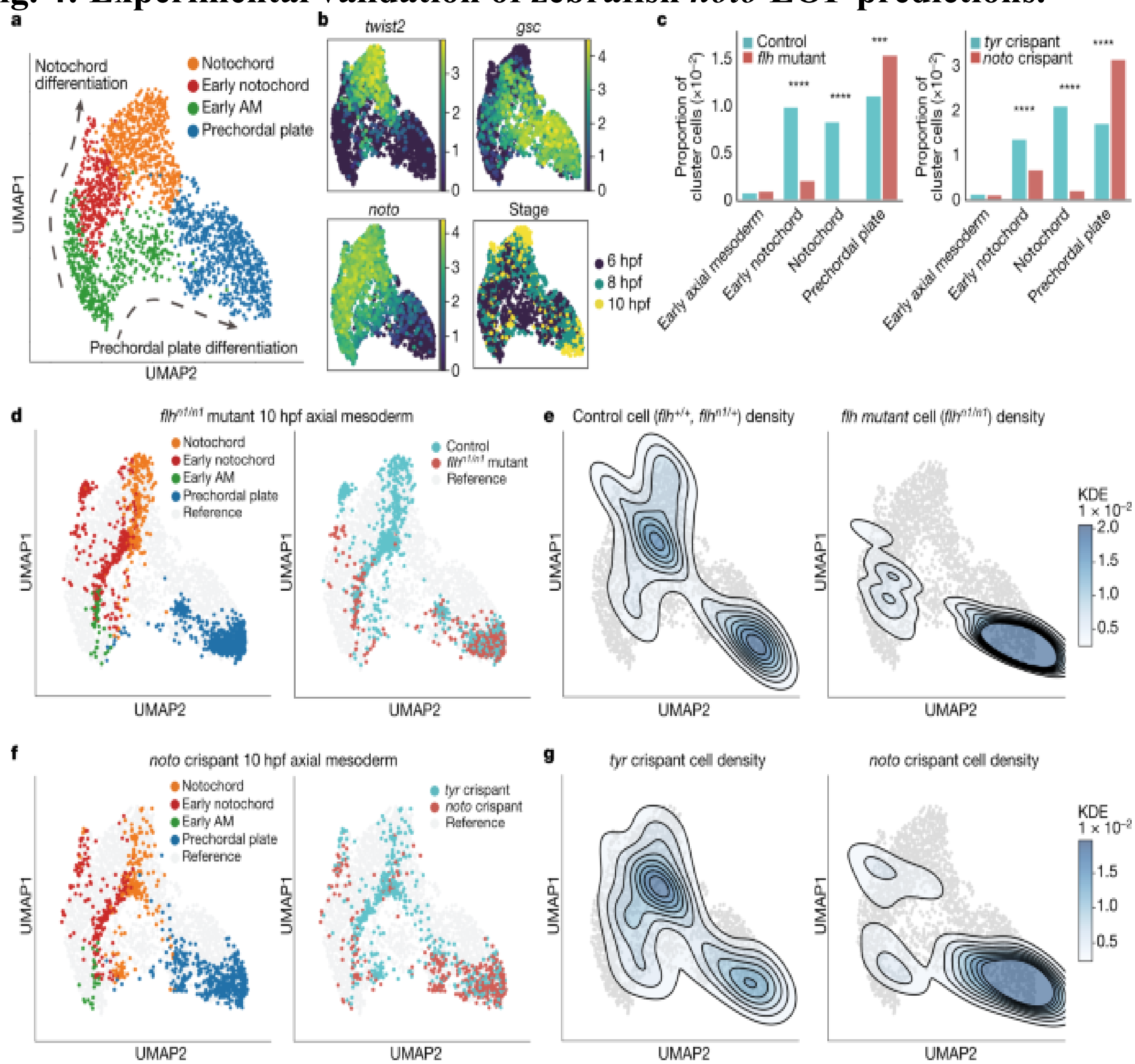


a, Two-dimensional force-directed graph of the axial mesoderm (AM) sub-branch ($n = 1,669$ cells) in a published zebrafish embryogenesis atlas (Farrell et al.[32]). Arrows indicate notochord cell differentiation (top) and prechordal plate differentiation (bottom). **b**, Conversion of URD-calculated pseudotime (left) into a 2D pseudotime gradient vector field (right). **c**, Degree centrality scores were used to rank the top 30 TFs in notochord (left) and prechordal plate (right). Black text denotes TFs. Grey text denotes non-TFs. **d**, Expression of *noto* projected onto the axial mesoderm sub-branch. **e**, *Noto* KO simulation vector and perturbation scores. **f**, Markov simulation to estimate cell density in the *noto* KO sample. The simulation predicted inhibited early notochord differentiation and promotion of prechordal plate differentiation, indicating a potential lineage switch.

Experimental validation of *noto* LOF

Next, we experimentally validated the predicted expansion of prechordal plate after *noto* LOF. First, we generated a 38,606-cell wild-type (WT) reference atlas from dissociated WT embryos at 6, 8 and 10 hpf (2 technical replicates per stage) and used Seurat's label transfer function[39] to cluster and label the WT reference cells according to the annotations in Farrell et al.[32] (Extended Data Fig. 11). Subsetting the axial mesoderm clusters showed the expected bifurcation of cells into notochord and prechordal plate, accompanied by upregulation of marker genes (Fig. 4a,b). For visualization of axial mesoderm cells, we used a uniform manifold approximation and projection (UMAP) transfer function to enable comparable data visualization between different samples (Methods).

Fig. 4: Experimental validation of zebrafish *noto* LOF predictions.



a, UMAP plot of WT reference data for axial mesoderm (6, 8 and 10 hpf): notochord, early notochord, early axial mesoderm and prechordal plate clusters ($n = 2,012$ cells). Arrows indicate notochord differentiation (top) and prechordal plate differentiation (bottom). **b**, Gene expression (log-transformed unique molecular identifier (UMI) count) and developmental stage are projected onto the axial mesoderm UMAP plot. *Noto* and *twist2* are expressed in notochord, whereas *gsc* marks the prechordal plate. **c**, Bar plots comparing cell cluster compositions between treatments and controls (left, $flh^{n1/n1}$ mutants (10 hpf) and controls; right, *noto* crispants (10 hpf) and *tyr* crispants). Cluster compositions are presented as the proportion of each group normalized to the whole cell number. In both $flh^{n1/n1}$ mutants and *noto* crispants, the notochord is significantly depleted ($flh^{n1/n1}$: $P = 5.55 \times 10^{-52}$; noto: $P = 1.39 \times 10^{-33}$, chi-square test) and the prechordal plate is significantly expanded ($flh^{n1/n1}$: $P = 1.07 \times 10^{-4}$; *noto*: $P = 5.01 \times 10^{-18}$, chi-square test. ***$P < 0.001$; ****$P < 0.0001$). **d**–**g**, $flh^{n1/n1}$ mutant or *noto* crispant data projected onto the WT axial mesoderm UMAP plot. **d**, Cluster annotation and sample label projected onto the UMAP plot. **e**, Kernel cell density contour plot shows control cell density (left) and $flh^{n1/n1}$ mutant cell density (right). **f**, Cluster annotation and sample label projected onto the UMAP plot. **g**, *tyr* crispant cell density (left) and *noto* crispant cell density (right) shown on the kernel cell density contour plot.

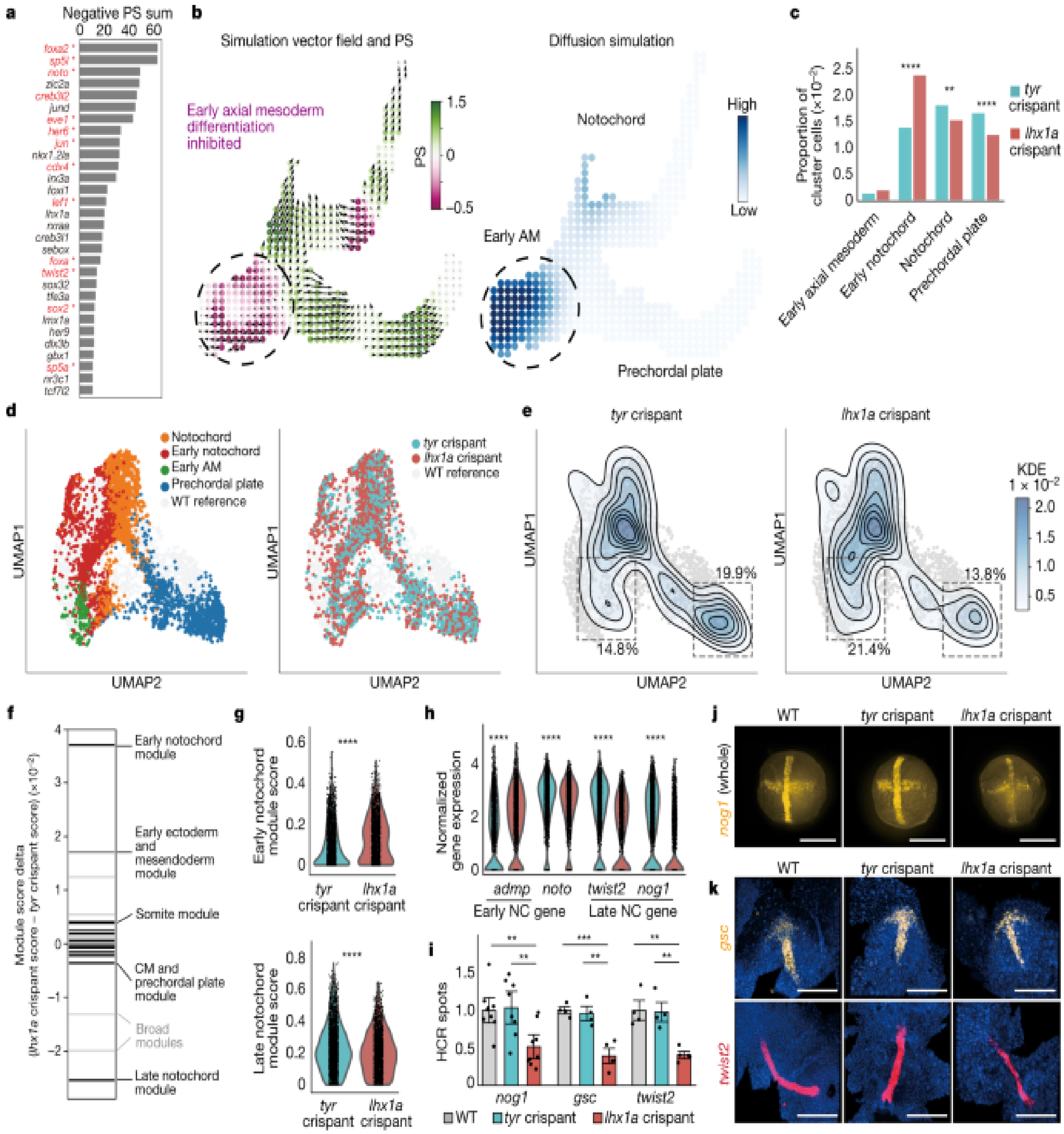
For experimental perturbation of *noto*, we generated and dissociated pools of 25 $flh^{n1/n1}$ mutant embryos, recognized at 10 hpf by the lack of notochord boundaries, and sibling controls ($flh^{n1/+}$ and $flh^{+/+}$) for scRNA-seq. We integrated these datasets and projected them onto the WT axial mesoderm reference atlas. In agreement with previous observations, we observed a significant depletion of cells labelled as notochord in $flh^{n1/n1}$ mutants (−98%, relative to control, $P = 5.55 \times 10^{-52}$, chi-square test; Fig. 4c, left), concomitant with an expansion of the somite cluster (+41.3%; $P = 5.90 \times 10^{-29}$; Extended Data Fig. 11e, left). Furthermore, as predicted by *noto* LOF simulation, we observed a significant expansion of the prechordal plate cluster in $flh^{n1/n1}$ mutants (+38.6%; $P = 1.07 \times 10^{-4}$; Fig. 4c, left). Plotting cell density revealed stalled notochord differentiation and bifurcation of the mid axial mesoderm, with excess prechordal plate cells (Fig. 4d,e), consistent with the *noto* LOF simulation (Fig. 3e,f). To orthogonally validate

these results, we produced *noto* LOF with a modified CRISPR–Cas9 protocol that we have previously used to achieve near-complete gene disruption in F_0 embryos injected with two *noto*-targeting guide RNAs (gRNAs)[40] (Methods). The resulting *noto* ‘crispants’ were dissociated at 10 hpf (9,185 cells, $n = 2$ biological and $n = 3$ technical replicates) and compared by single-cell analysis to controls that targeted the *tyrosinase* gene (*tyr)*, which is not expressed until later in development ($n = 46{,}440$ single cells, $n = 3$ biological and $n = 5$ technical replicates; Extended Data Fig. 11b). Analysis of cell-type composition confirmed a significant depletion of notochord, with an expansion of somitic mesoderm and prechordal plate (Fig. 4c, right, Fig. 4f,g and Extended Data Fig. 11e, right) in *noto* crispants, highly consistent with our $flh^{n1/n1}$ mutant analysis. Together, in addition to further validating the performance of CellOracle, these results highlight the ability of this approach to identify experimentally quantifiable phenotypes in well-characterized mutants that may have been previously overlooked owing to a reliance on gross morphology. We next sought to identify new LOF phenotypes in axial mesoderm development.

Discovery of axial mesoderm regulators

To identify novel TFs required for axial mesoderm differentiation, we prioritized TFs according to predicted KO phenotypes, focusing on early-stage differentiation before evident lineage specification (Extended Data Fig. 12a). The resulting ranked list contains several known notochord regulators, including *noto* (Fig. 5a, red and Supplementary Table 2), confirming CellOracle’s capacity to model known developmental regulation. However, it is important to note that some known notochord regulators, such as *foxa3* (ref. [41]), were not identified as they are filtered out in the first steps of data processing owing to low expression. Systematic perturbation simulations for all lineages can be found at https://celloracle.org. As well as the axial mesoderm, we also performed an in-depth analysis of the adaxial mesoderm, which gives rise to somites. Overall, more than 80% of the top 30 TFs in this analysis were associated with somite differentiation (Supplementary Table 3).

Fig. 5: Experimental validation of *lhx1a* as a putative regulator of zebrafish axial mesoderm development.



a, Top 30 TFs according to predicted KO effects. Red and *: previously reported notochord regulators (Supplementary Table 2). *lhx1a*, *sebox* and *irx3a* were selected for experimental validation. **b**, *lhx1a* LOF simulation in the axial mesoderm sub-branch, predicting an inhibition of axial mesoderm differentiation from early stages. **c**, scRNA-seq validation of experimental LOF: cell cluster composition of the axial mesoderm clusters normalized to the whole cell number in *lhx1a* and *tyr* (control) crispant samples. Early notochord is significantly expanded ($P = 1.34 \times 10^{-35}$, chi-square test) and differentiated axial mesoderm populations are significantly depleted

(notochord: $P = 3.83 \times 10^{-3}$; prechordal plate: $P = 1.28 \times 10^{-7}$, chi-square test) in *lhx1a* crispants. **d**, *lhx1a* and *tyr* crispant axial mesoderm cells at 10 hpf. Left, cell type annotation of *lhx1* and *tyr* crispant cells. Right, *lhx1a* and control crispant data projected onto the WT UMAP. **e**, Control cell density (left, $n = 2{,}342$ cells) and *lhx1a* crispant cell density (right, $n = 2{,}502$ cells). **f**, Rug plot showing the difference in averaged NMF module scores between *lhx1a* and *tyr* crispants in notochord lineage cells. Black, cell-type-specific modules. Light grey, broad cluster modules. CM, cephalic mesoderm. **g**, Violin plot of NMF module score in notochord lineage cells ($n = 1{,}918$ *lhx1a* crispant and $n = 2{,}616$ *tyr* crispant cells. **h**, Violin plots of gene expression in the notochord (NC) lineage cells. $^{****}P < 0.0001$, two-tailed Wilcoxon rank-sum test with Bonferroni correction. **i**, Quantification (number of spots in flattened HCR image) normalized to WT. Mean ± s.e.m. $n = 2$ independent biological replicates, 8 embryos per replicate. *nog1*: $P = 0.0022$ (WT versus *lhx1a* crispant), $P = 0.0052$ (*tyr* versus *lhx1a* crispant); *gsc*: $P = 0.00042$ (WT versus *lhx1a* crispant), $P = 0.0018$ (*tyr* versus *lhx1a* crispant); *twist2*: $P = 0.0011$ (WT versus *lhx1a* crispant), $P = 0.0012$ (*tyr* versus *lhx1a* crispant); two-sided *t*-test. **j**, Representative HCR images for *nog1* expression (yellow) in whole embryos at 10 hpf. **k**, Representative flattened HCR images of 10 hpf embryos stained with probes against *gsc* (yellow) and *twist2* (red); nuclei are stained with DAPI (blue). Scale bars, 300 μm.

In addition to known TFs, we identified several TFs with no previously reported role in axial mesoderm differentiation (Fig. 5a, black). We further prioritized candidate genes for experimental validation by GRN degree centrality, gene enrichment score in axial mesoderm and average gene expression value, selecting *lhx1a*, *sebox* and *irx3a* (Extended Data Fig. 12b). CellOracle predicts impaired notochord differentiation for all three genes after their LOF (Fig. 5b and Supplementary Fig. 19). However, no LOF studies describing axial mesoderm phenotypes that relate to these genes have, to our knowledge, been reported in zebrafish. Mouse *Lhx1 (Lim1)* KO embryos lack anterior head structures and kidneys[42]. In zebrafish, *sebox* (*mezzo*) has been implicated in mesoderm and endoderm specification[43], whereas *irx3a* (*ziro3*) morphants exhibit changes in the composition of pancreatic cell types[44].

We generated *lhx1a*, *sebox* and *irx3a* crispants (Supplementary Fig. 20b–d). We performed initial single-cell analyses at 10 hpf, integrating crispant scRNA-seq datasets with the control gRNA reference atlas described above. We observed significant changes in cell-type composition and notochord marker expression in *lhx1a* crispants (Extended Data Fig. 12c,d and Supplementary Table 4). Notably, we found a more considerable reduction in the expression of late notochord genes relative to broad notochord markers, suggesting that loss of *lhx1a* function inhibits the differentiation and maturation of notochord cells. We observed a slight yet significant reduction in the expression of the notochord markers *twist2*, *nog1* and *tbxta* in *sebox* crispants (Extended Data Fig. 12e,f and Supplementary Table 4), confirming CellOracle's predictions that *lhx1a* and *sebox* are regulators of axial mesoderm development. *Irx3a* crispants showed no significant phenotype in cell-type composition but exhibited a slight reduction in *twist2* expression in the notochord (Extended Data Fig. 12g,h).

We extended *lhx1a* LOF characterization by performing four independent biological replicates for *lhx1a* crispants ($n = 45{,}582$ cells) and *tyr* crispants ($n = 76{,}163$ cells, 5 biological and 7 technical replicates). CellOracle predicted inhibition of early axial mesoderm differentiation after *lhx1a* disruption, depleting both notochord and prechordal plate lineages (Fig. 5b). Indeed, the *lhx1a* crispants exhibited inhibition of axial mesoderm differentiation (Fig. 5c–e): a significant expansion of the early notochord cluster (+70.2%; $P = 1.34 \times 10^{-35}$), with a concomitant reduction of later notochord (−15.3%; $P = 3.83 \times 10^{-3}$) and prechordal plate clusters (−24.7%; $P = 1.28 \times 10^{-7}$). These phenotypes were reproducible across independent biological replicates (Extended Data Fig. 13e), validating the predicted inhibition of early axial mesoderm differentiation (Fig. 5a,b).

To further analyse the *lhx1a* LOF axial mesoderm phenotype, we investigated global changes in gene expression across all cell types using non-negative matrix factorization (NMF), a method to quantify gene module activation[45] (Supplementary Table 5 and Methods). We observed that a module corresponding to the early notochord was significantly activated in *lhx1a* crispants ($P = 2.62 \times 10^{-32}$; Fig. 5f,g). The top gene in this module is *admp* (Extended Data Fig. 13f, left), which is significantly upregulated in *lhx1a* crispant cells ($P = 6.69 \times 10^{-46}$; Fig. 5h) and encodes a known

negative regulator of notochord and prechordal plate development[46]. By contrast, the late notochord module received a significantly lower score in the *lhx1a* crispant cells ($P = 1.04 \times 10^{-5}$; Fig. 5g, bottom). This module comprises late notochord marker genes, such as *twist2* and *nog1* (Extended Data Fig. 13f, right), which showed significantly lower expression in *lhx1a* crispant cells ($P = 4.52 \times 10^{-105}$ and $P = 4.95 \times 10^{-105}$, respectively; Fig. 5h). Further, *lhx1a* crispant cells exhibited a higher somite module score ($P = 5.19 \times 10^{-25}$ and Supplementary Table 5), suggesting that notochord cells may be redirected towards a somitic identity after *lhx1a* LOF. Overall, the NMF analysis supports the hypothesis that loss of *lhx1a* function induces global changes in gene expression that are related to inhibited notochord differentiation.

Finally, we confirmed the *lhx1a* LOF phenotype using orthogonal approaches. Hybridization chain reaction (HCR) RNA fluorescence in situ hybridization for *nog1* (late notochord) and for *gsc* and *twist2* (prechordal plate and notochord, respectively) showed that these genes were significantly downregulated in *lhx1a* crispants (Fig. 5i–k). These results were further confirmed by quantitative reverse transcription PCR (qRT-PCR) and whole-mount in situ hybridization against *nog1* (Supplementary Fig. 22). Together, this experimental validation confirms the significant and consistent disruption of axial mesoderm development after loss of *lhx1a* function. In summary, these results demonstrate the ability of CellOracle to accurately predict known TF perturbation phenotypes, provide insight into previously characterized mutants and reveal regulators of established developmental processes in well-studied model organisms.

Discussion

The emerging discipline of perturbational single-cell omics enables regulators of cell identity and behaviour to be modelled and predicted[5]. For example, scGen combines variational autoencoders with latent space vector arithmetic to predict cell infection response. However, this approach requires experimentally perturbed training data, which limits its scalability[47]. More importantly, it remains challenging to interpret the gene program behind the simulated outcome using these previous computational perturbation approaches because they rely on complex black-box models; thus, the

simulations lack any means to interpret how gene regulation relates to cellular phenotype. On the other hand, previous GRN analyses relied largely on static graph theory and could not consider cell identity as a dynamic property. Here we present a strategy that overcomes these limitations by integrating computational perturbation with GRN modelling. CellOracle uses GRN models to yield mechanistic insights into the regulation of cell identity; simulation and vector visualization based on the custom network model enables the interpretable, scalable and broadly applicable analysis of dynamic TF function.

We validated CellOracle using various in vivo differentiation models, verifying its efficacy and its robustness to complex and noisy biological data. CellOracle simulates shifts in cell identity by considering systematic gene-to-gene relationships for each cell state using multimodal data, generating a complex context-dependent vector representation that is not possible using differential gene expression or chromatin accessibility alone. For example, the role of *Gata1* in granulocyte differentiation would probably not be predicted given its low expression in this cell type. However, CellOracle could corroborate this relatively mild *Gata1* phenotype. Furthermore, CellOracle correctly reproduced the reported early-stage-specific cell-fate-regulatory role of *Tal1* in erythropoiesis, which is impossible to uncover on the basis of the constitutive expression of *Tal1* throughout all erythroid stages. This capacity of CellOracle means that it could identify previously unreported phenotypes. For example, the LOF simulation of a well-characterized regulator of zebrafish axial mesoderm development, *noto*, predicted a previously unreported expansion of the prechordal plate, which we experimentally validated. This case suggests that *noto* has a role in suppressing alternate fates, which could only be predicted by the integrative simulation using the GRN and cell differentiation trajectory together. Finally, although we focus on TF KO and LOF in this study, we have also recently demonstrated that CellOracle can be used to simulate TF overexpression[48].

We note some limitations of the method. First, CellOracle visualizes the simulation vector within the existing trajectory space; thus, cell states that do not exist in the input scRNA-seq data cannot be analysed. Nevertheless, existing single-cell data collected after severe developmental disruption do not report the emergence of new transcriptional states in the context of loss

of gene function, which suggests extensive canalization even during abnormal development[32], supporting the use of CellOracle to accurately simulate TF perturbation effects. Second, we emphasize that TF simulation is limited by input data availability and data quality. For example, a perturbation cannot be simulated if a TF-binding motif is unknown or TF expression is too sparse, as we note in the case of *foxa3* in zebrafish[41].

Our application of CellOracle to systematically simulate TF perturbation has revealed regulators of a well-characterized developmental paradigm: the formation of axial mesoderm in zebrafish. Although zebrafish axial mesoderm has been well-characterized through mutagenesis screens, a role for Lhx1a in these developmental stages is likely to have gone unreported owing to the absence of gross morphological phenotypical changes at 10 hpf after disruption of *lhx1a* (ref. [49]). However, our ability to predict and validate such a phenotype showcases the power of single-cell computational and experimental approaches, enabling finer-resolution dissection of gene regulation even in well-characterized systems. Moreover, CellOracle provides information at intermediate steps in a given developmental pathway, obviating the need for gross morphological end-points. Indeed, each simulation can be thought of as many successive predictions along a lineage, although we stress that experimental validation is essential to validate CellOracle's predictions where possible. However, applying these approaches to emerging systems or where experimental intervention is not feasible promises to accelerate our understanding of how cell identity is regulated. For example, in the context of human development, we have recently applied CellOracle to predict candidate regulators of medium spiny neuron maturation in human fetal striatum[50], demonstrating the power of in silico perturbation where experimental approaches cannot be deployed.

Methods

CellOracle algorithm overview

The CellOracle workflow consists of several steps: (1) base GRN construction using scATAC-seq data or promoter databases; (2) scRNA-seq data preprocessing; (3) context-dependent GRN inference using scRNA-seq data; (4) network analysis; (5) simulation of cell identity following TF

perturbation; and (6) calculation of the pseudotime gradient vector field and the inner-product score to generate perturbation scores. We implemented and tested CellOracle in Python (versions 3.6 and 3.8) and designed it for use in the Jupyter notebook environment. CellOracle code is open source and available on GitHub (https://github.com/morris-lab/CellOracle), along with detailed descriptions of functions and tutorials.

Base GRN construction using scATAC-seq data

In the first step, CellOracle constructs a base GRN that contains unweighted, directional edges between a TF and its target gene. CellOracle uses the regulatory region's genomic DNA sequence and TF-binding motifs for this task. CellOracle identifies regulatory candidate genes by scanning for TF-binding motifs within the regulatory DNA sequences (promoter and enhancers) of open chromatin sites. This process is beneficial as it narrows the scope of possible regulatory candidate genes in advance of model fitting and helps to define the directionality of regulatory edges in the GRN. However, the base network generated in this step may still contain pseudo- or inactive connections; TF regulatory mechanisms are not only determined by the accessibility of binding motifs but may also be influenced by many context-dependent factors. Thus, scRNA-seq data are used to refine this base network during the model fitting process in the next step of base GRN assembly.

Base GRN assembly can be divided into two steps: (i) identification of promoter and enhancer regions using scATAC-seq data; and (ii) motif scanning of promoter and enhancer DNA sequences.

Identification of promoter and enhancer regions using scATAC-seq data

CellOracle uses genomic DNA sequence information to define candidate regulatory interactions. To achieve this, the genomic regions of promoters and enhancers first need to be designated, which we infer from ATAC-seq data. We designed CellOracle for use with scATAC-seq data to identify accessible promoters and enhancers (Extended Data Fig. 1a, left panel). Thus, scATAC-seq data for a specific tissue or cell type yield a base GRN representing a sample-specific TF-binding network. In the absence of a

sample-specific scATAC-seq dataset, we recommend using scATAC-seq data from closely related tissue or cell types to support the identification of promoter and enhancer regions. Using broader scATAC-seq datasets produces a base GRN corresponding to a general TF-binding network rather than a sample-specific base GRN. Nevertheless, this base GRN network will still be tailored to a specific sample using scRNA-seq data during the model fitting process. The final product will consist of context-dependent (cell-type or state-specific) GRN configurations.

To identify promoter and enhancer DNA regions within the scATAC-seq data, CellOracle first identifies proximal regulatory DNA elements by locating TSSs within the accessible ATAC-seq peaks. This annotation is performed using HOMER (http://homer.ucsd.edu/homer/). Next, the distal regulatory DNA elements are obtained using Cicero, a computational tool that identifies *cis*-regulatory DNA interactions on the basis of co-accessibility, as derived from ATAC-seq peak information[12]. Using the default parameters of Cicero, we identify pairs of peaks within 500 kb of each other and calculate a co-accessibility score. Using these scores as input, CellOracle then identifies distal *cis*-regulatory elements defined as pairs of peaks with a high co-accessibility score (≥0.8), with the peaks overlapping a TSS. The output is a bed file in which all *cis*-regulatory peaks are paired with the target gene name. This bed file is used in the next step. CellOracle can also use other input data types to define *cis*-regulatory elements. For example, a database of promoter and enhancer DNA sequences or bulk ATAC-seq data can serve as an alternative if available as a .bed file.

For the analysis of mouse haematopoiesis that we present here, we assembled the base GRN using a published mouse scATAC-seq atlas consisting of around 100,000 cells across 13 tissues, representing around 400,000 differentially accessible elements and 85 different chromatin patterns[13]. This base GRN is built into the CellOracle library to support GRN inference without sample-specific scATAC-seq datasets. In addition, we have generated general promoter base GRNs for several key organisms commonly used to study development, including 10 species and 23 reference genomes (Supplementary Table 1).

Motif scan of promoter and enhancer DNA sequences

This step scans the DNA sequences of promoter and enhancer elements to identify TF-binding motifs. CellOracle internally uses gimmemotifs (https://gimmemotifs.readthedocs.io/en/master/), a Python package for TF motif analysis. For each DNA sequence in the bed file obtained in step (i) above, motif scanning is performed to search for TF-binding motifs in the input motif database.

For mouse and human data, we use gimmemotifs motif v.5 data. CellOracle also provides a motif dataset for ten species generated from the CisBP v.2 database (http://cisbp.ccbr.utoronto.ca).

CellOracle exports a binary data table representing a potential connection between a TF and its target gene across all TFs and target genes. CellOracle also reports the TF-binding DNA region. CellOracle provides pre-built base GRNs for ten species (Supplementary Table 1), which can be used if scATAC-seq data are unavailable.

scRNA-seq data preprocessing

CellOracle requires standard scRNA-seq preprocessing in advance of GRN construction and simulation. The scRNA-seq data need to be prepared in the AnnData format (https://anndata.readthedocs.io/en/latest/). For data preprocessing, we recommend using Scanpy (https://scanpy.readthedocs.io/en/stable/) or Seurat (https://satijalab.org/seurat/). Seurat data must be converted into the AnnData format using the CellOracle function, seuratToAnndata, preserving its contents. In the default CellOracle scRNA-seq preprocessing step, zero-count genes are first filtered out by UMI count using scanpy.pp.filter_genes(min_counts=1). After normalization by total UMI count per cell using sc.pp.normalize_per_cell(key_n_counts='n_counts_all'), highly variable genes are detected by scanpy.pp.filter_genes_dispersion(n_top_genes=2000~3000). The detected variable gene set is used for downstream analysis. Gene expression values are log-transformed, scaled and subjected to dimensional reduction and clustering. The non-log-transformed gene expression matrix (GEM) is also retained, as it is required for downstream GRN calculation and simulation.

Context-dependent GRN inference using scRNA-seq data

In this step of CellOracle GRN inference, a machine-learning model is built to predict target gene expression from the expression levels of the regulatory genes identified in the previous base GRN refinement step. By fitting models to sample gene expression data, CellOracle extracts quantitative gene–gene connection information. For signal propagation, the CellOracle GRN model must meet two requirements: (1) the GRN model needs to represent transcriptional connections as a directed network edge; and (2) the GRN edges need to be a linear regression model. Because of this second constraint, we cannot use pre-existing GRN inference algorithms, such as GENIE3 and GRNboost (refs. [7,51]). CellOracle leverages genomic sequences and information on TF-binding motifs to infer the base GRN structure and directionality, and it does not need to infer the causality or directionality of the GRN from gene expression data. This allows CellOracle to adopt a relatively simple machine-learning model for GRN inference—a regularized linear machine-learning model. CellOracle builds a model that predicts the expression of a target gene on the basis of the expression of regulatory candidate genes:

$${x}_{j}=\,\mathop{\sum }\limits_{i=0}^{n}{b}_{i,j}{x}_{i}+\,{c}_{j},$$

where x_j is single target gene expression and x_i is the gene expression value of the regulatory candidate gene that regulates gene x_j. $b_{i,j}$ is the coefficient value of the linear model (but $b_{i,j} = 0$ if $i = j$), and c is the intercept for this model. Here, we use the list of potential regulatory genes for each target gene generated in the previous base GRN construction step (ii).

$${x}_{i}\in \{{x}_{0},\,{x}_{1},\,\ldots {x}_{n}\}={\rm{Regulatory}}\,{\rm{candidate\; TFs\; of\; gene}}\,{x}_{j}$$

The regression calculation is performed for each cell cluster in parallel after the GEM of scRNA-seq data is divided into several clusters. The cluster-wise regression model can capture non-linear or mixed regulatory relationships. In addition, L2 weight regularization is applied by the Ridge model. Regularization not only helps distinguish active regulatory

connections from random, inactive, or false connections in the base GRN but also reduces overfitting in smaller samples.

The Bayesian Ridge or Bagging Ridge model provides the coefficient value as a distribution, and we can analyse the reproducibility of the inferred gene–gene connection (Extended Data Fig. 1a, right). In both models, the output is a posterior distribution of coefficient value *b*:

$${x}_{j}\sim {\rm{N}}{\rm{o}}{\rm{r}}{\rm{m}}{\rm{a}}{\rm{l}}\,\,(\mathop{\sum }\limits_{i=1}^{n}{b}_{i,j}{x}_{i}+{c}_{j},{\epsilon })$$

$$b\sim {\rm{N}}{\rm{o}}{\rm{r}}{\rm{m}}{\rm{a}}{\rm{l}}\,(\,{\mu }_{b},{{\sigma }}_{b})$$

where ${\mu }_{b}$ is the centre of the distribution of *b*, and ${\sigma }_{b}$ is the standard deviation of *b*. The user can choose the model method depending on the availability of computational resources and the aim of the analysis; CellOracle's Bayesian Ridge requires fewer computational resources, whereas the Bagging Ridge tends to produce better inference results than Bayesian Ridge. Using the posterior distribution, we can calculate *P* values of coefficient *b*; one-sample *t*-tests are applied to *b* to estimate the probability (the centre of $b = 0$). The *P* value helps to identify robust connections while minimizing connections derived from random noise. In addition, we apply regularization to coefficient *b* for two purposes: (i) to prevent coefficient *b* from becoming extremely large owing to overfitting; and (ii) to identify informative variables through regularization. In CellOracle, the Bayesian Ridge model uses regularizing prior distribution of *b* as follows:

$$b\sim {\rm{N}}{\rm{o}}{\rm{r}}{\rm{m}}{\rm{a}}{\rm{l}}\,(0,{{\sigma }}_{b})$$

$${{\sigma }}_{b}^{-1}\sim {\rm{G}}{\rm{a}}{\rm{m}}{\rm{m}}{\rm{a}}\,({10}^{-6},{10}^{-6})$$

${\sigma }_{b}$ is selected to represent non-informative prior distributions. This model uses data in the fitting process to estimate the optimal regularization strength. In the Bagging Ridge model, custom regularization strength can be manually set.

For the computational implementation of the above machine-learning models, we use a Python library, scikit-learn (https://scikit-learn.org/stable/). For Bagging Ridge regression, we use the Ridge class in the sklearn.linear_model and BaggingRegressor in the sklearn.ensemble module. The number of iterative calculations in the bagging model can be adjusted depending on the computational resources and available time. For Bayesian Ridge regression, we use the BayesianRidge class in sklearn.linear_module with the default parameters.

Simulation of cell identity following perturbation of regulatory genes

The central purpose of CellOracle is to understand how a GRN governs cell identity. Toward this goal, we designed CellOracle to make use of inferred GRN configurations to simulate how cell identity changes following perturbation of regulatory genes. The simulated gene expression values are converted into 2D vectors representing the direction of cell-state transition, adapting the visualization method previously used by RNA velocity[52]. This process consists of four steps: (i) data preprocessing; (ii) signal propagation within the GRN; (iii) estimation of transition probabilities; and (iv) analysis of simulated transition in cell identity.

1. (i)

   Data preprocessing

   For simulation of cell identity, we developed our code by modifying Velocyto.py, a Python package for RNA-velocity analysis (https://velocyto.org). Consequently, CellOracle preprocesses the scRNA-seq data per Velocyto requirements by first filtering the genes and imputing dropout. Dropout can affect Velocyto's transition probability calculations; thus, *k*-nearest neighbour (KNN) imputation must be performed before the simulation step.

2. (ii)

   Within-network signal propagation

This step aims to estimate the effect of TF perturbation on cell identity. CellOracle simulates how a 'shift' in input TF expression leads to a 'shift' in its target gene expression and uses a partial derivative $\frac{\partial {x}_{j}}{\partial {x}_{i}}$. As we use a linear model, the derivative $\frac{\partial {x}_{j}}{\partial {x}_{i}}$ is a constant value and already calculated as $b_{i,j}$ in the previous step if the gene *j* is directly regulated by gene *i*:

$$\frac{\partial {x}_{j}}{\partial {x}_{i}}={b}_{i,j}.$$

And we calculate the shift of target gene ${\Delta x}_{j}$ in response to the shift of regulatory gene ${\Delta x}_{i}$:

$${\Delta x}_{j}=\frac{\partial {x}_{j}}{\partial {x}_{i}}{\Delta x}_{i}={b}_{i,j}{\Delta x}_{i}.$$

As we want to consider the gene-regulatory 'network', we also consider indirect connections. The network edge represents a differentiable linear function shown above, and the network edge connections between indirectly connected nodes is a composite function of the linear models, which is differentiable accordingly. Using this feature, we can apply the chain rule to calculate the partial derivative of the target genes, even between indirectly connected nodes.

$$\frac{\partial {x}_{j}}{\partial {x}_{i}}=\mathop{\prod }\limits_{k=0}^{n}\frac{\partial {x}_{k+1}}{\partial {x}_{k}}=\mathop{\prod }\limits_{k=0}^{n}{b}_{k,k+1},$$

where

$$\begin{array}{c}{x}_{k}\in \{{x}_{0},\,{x}_{1},\,\ldots {x}_{n}\}\,=\,\text{Gene expression of ordered network}\\ \,\text{nodes on the shortest path from gene}\,i\,\text{to gene}\,j.\end{array}$$

For example, when we consider the network edge from gene 0 to 1 to 2, the small shift of gene 2 in response to gene 0 can be calculated using the intermediate connection with gene 1 (Supplementary Fig. 1).

$$\frac{\partial {x}_{2}}{\partial {x}_{0}}=\frac{\partial {x}_{1}}{\partial {x}_{0}}\times \frac{\partial {x}_{2}}{\partial {x}_{1}}={b}_{0,1}\times {b}_{1,2}$$

$${\Delta x}_{2}=\frac{\partial {x}_{2}}{\partial {x}_{0}}{\Delta x}_{0}={b}_{0,1}{b}_{1,2}{\Delta x}_{0}$$

In summary, the small shift of the target gene can be formulated by the multiplication of only two components, GRN model coefficient $b_{i,j}$ and input TF shift $({\Delta x}_{i})$. In this respect, we focus on the gradient of gene expression equations rather than the absolute expression values so that we do not model the error or the intercept of the model, which potentially includes unobservable factors within the scRNA-seq data.

The calculation above is implemented as vector and matrix multiplication. First, the linear regression model can be shown as follows.

$${X}^{{\prime} }=X\cdot B+C,$$

where the $X\in {{\mathbb{R}}}^{1\times N}$ is a gene expression vector containing N genes, $C\in {{\mathbb{R}}}^{1\times N}$ is the intercept vector, $B\in {{\mathbb{R}}}^{N\times N}$ is the network adjacency matrix, and each element $b_{i,j}$ is the coefficient value of the linear model from regulatory gene i to target gene j.

First, we set the perturbation input vector $\Delta {X}_{{\rm{input}}}\in {{\mathbb{R}}}^{1\times N}$, a sparse vector consisting of zero except for the perturbation target gene i. For the TF perturbation target gene, we set the shift of the TF to be simulated. The CellOracle function will produce an error if the user enters a gene shift corresponding to an out-of-distribution value.

Next, we calculate the shift of the first target gene:

$$\Delta {X}_{{\rm{simulated}},n=1}=\Delta {X}_{{\rm{input}}}\cdot B.$$

However, we fix the perturbation target gene *i* value, and the ${\Delta x}_{i}$ retains the same value as the input state. Thus, the following calculation will correspond to both the first and the second downstream gene shift calculations.

$$\Delta {X}_{{\rm{simulated}},n=2}=\Delta {X}_{{\rm{simulated}},n=1}\cdot B.$$

Likewise, the recurrent calculation is performed to propagate the shift from gene to gene in the network. Repeating this calculation for *n* iterations, we can estimate the effects on the first to the *n*th indirect target gene (Extended Data Fig. 1b–d):

$$\Delta {X}_{{\rm{simulated}},n}=\Delta {X}_{{\rm{simulated}},n-1}\cdot B.$$

CellOracle performs three iterative cycles in the default setting, sufficient to predict the directionality of changes in cell identity (Supplementary Figs. 4 and 5). We avoid a higher number of iterative calculations as it might lead to unexpected behaviour. Of note, CellOracle performs the calculations cluster-wise after splitting the whole GEM into gene expression submatrices on the basis of the assumption that each cluster has a unique GRN configuration. Also, gene expression values are checked between each iterative calculation to confirm whether the simulated shift corresponds to a biologically plausible range. If the expression value for a gene is negative, this value is adjusted to zero. The code in this step is implemented from scratch, specifically for CellOracle perturbations using NumPy, a python package for numerical computing (https://numpy.org).

3. (iii)

   Estimation of transition probabilities

   From the previous steps, CellOracle produces a simulated gene expression shift vector $\Delta {X}_{{\rm{simulated}}}\in {{\mathbb{R}}}^{1\times N}$ representing the simulated initial gene expression shift after TF perturbation. Next, CellOracle aims to project

the directionality of the future transition in cell identity onto the dimensional reduction embedding (Fig. 1a, right and Extended Data Fig. 1e). For this task, CellOracle uses a similar approach to Velocyto (https://github.com/velocyto-team/velocyto.py). Velocyto visualizes future cell identity on the basis of the RNA-splicing information and calculated vectors from RNA synthesis and degradation differential equations. CellOracle uses the simulated gene expression vector $\Delta {X}_{{\rm{simulated}}}$ instead of RNA-velocity vectors.

First, CellOracle estimates the cell transition probability matrix $P\in {{\mathbb{R}}}^{M\times M}$ (*M* is number of cells): $p_{i,j}$, the element in the matrix *P*, is defined as the probability that cell *i* will adopt a similar cell identity to cell *j* after perturbation. To calculate $p_{i,j}$, CellOracle calculates the Pearson's correlation coefficient between d_i and $r_{i,j}$:

$${p}_{ij}=\frac{\exp \left(corr\left({r}_{ij}{,d}_{i}\right)/T\right)}{\sum _{j\in G}\exp \left(corr\left({r}_{ij}{,d}_{i}\right)/T\right)},$$

where d_i is the simulated gene expression shift vector $\Delta {X}_{{\rm{simulated}}}\in {{\mathbb{R}}}^{1\times N}$ for cell *i*, and ${r}_{ij}\in {{\mathbb{R}}}^{1\times N}$ is a subtraction of the gene expression vector $X\in {{\mathbb{R}}}^{1\times N}$ between cell *i* and cell *j* in the original GEM. The value is normalized by the Softmax function (default temperature parameter *T* is 0.05). The calculation of $p_{i.j}$ uses neighbouring cells of cell i. The KNN method selects local neighbours in the dimensional reduction embedding space ($k = 200$ as default).

4. (iv)

   Calculation of simulated cell-state transition vector

   The transition probability matrix *P* is converted into a transition vector ${V}_{i,{\rm{simulated}}}\in {{\mathbb{R}}}^{1\times 2}$, representing the relative cell-identity shift of cell *i* in the 2D dimensional reduction space, as follows: CellOracle calculates the local

weighted average of vector ${V}_{i,j}\in {{\mathbb{R}}}^{1\times 2},{V}_{i,j}$ denotes the 2D vector obtained by subtracting the 2D coordinates in the dimensional reduction embedding between cell *i* and cell *j* (${\rm{cell}}\;j\in G$).

$${V}_{i,{\rm{simulated}}}=\,\sum _{j\in G}{p}_{ij}{V}_{i,j}$$

5. (v)

Calculation of vector field

The single-cell resolution vector ${V}_{i,{\rm{simulated}}}$ is too fine to interpret the results in a large dataset consisting of many cells. We calculate the summarized vector field using the same vector averaging strategy as Velocyto. The simulated cell-state transition vector for each cell is grouped by grid point to get the vector field, ${V}_{{\rm{vector}}{\rm{field}}}=\,{{\mathbb{R}}}^{2\times L\times L}$, (*L* is grid number, default *L* is 40). ${v}_{{\rm{grid}}}\in {{\mathbb{R}}}^{2}$, an element in the ${V}_{{\rm{vector}}{\rm{field}}}$, is calculated by the Gaussian kernel smoothing.

$${v}_{{\rm{grid}}}={\sum }_{i\in H}{K}_{\sigma }(g,\,{V}_{i,{\rm{simulated}}}){V}_{i,{\rm{simulated}}},$$

where the $g\in {{\mathbb{R}}}^{2}$ denotes grid point coordinates, *H* is the neighbour cells of *g* and ${K}_{\sigma }$ is the Gaussian kernel weight:

$${K}_{\sigma }({v}_{0},{v}_{1})=\exp \left(\frac{-{\parallel {v}_{0}-{v}_{1}\parallel }^{2}}{2{\sigma }^{2}}\right).$$

Calculation of pseudotime gradient vector field and inner-product score to generate a perturbation score

To aid the interpretation of CellOracle simulation results, we quantify the similarity between the differentiation vector fields and KO simulation vector fields by calculating their inner-product value, which we term the perturbation score (PS) (Extended Data Fig. 4). Calculation of the PS includes the following steps:

1. (i)

   Differentiation pseudotime calculation

   Differentiation pseudotime is calculated using DPT, a diffusion-map-based pseudotime calculation algorithm, using the scanpy.tl.dpt function (Extended Data Fig. 4a, left). CellOracle also works with other pseudotime data, such as Monocle pseudotime and URD pseudotime data. For the Farrell et al.[32] zebrafish scRNA-seq data analysis, we used pseudotime data calculated by the URD algorithm, as described previously[32].

2. (ii)

   Differentiation vector calculation based on pseudotime data

   The pseudotime data are transferred to the *n* by *n* 2D grid points (n = 40 as default) (Extended Data Fig. 4a, centre). For this calculation, we implemented two functions in CellOracle: KNN regression and polynomial regression for the data transfer. We choose polynomial regression when the developmental branch is a relatively simple bifurcation, as is the case for the Paul et al.[16] haematopoiesis data. We used KNN regression for a more complex branching structure, such as the Farrell et al.[32] zebrafish development data. Then, CellOracle calculates the gradient of pseudotime data on the 2D grid points using the numpy.gradient function, producing the 2D vector map

representing the direction of differentiation (Extended Data Fig. 4a, right).

3. (iii)

   Inner-product value calculation between differentiation and KO simulation vector field

   Then, CellOracle calculates the inner-product score (perturbation score (PS)) between the pseudotime gradient vector field and the perturbation simulation vector field (Extended Data Fig. 4b). The inner product between the two vectors represents their agreement (Extended Data Fig. 4c), enabling a quantitative comparison of the directionality of the perturbation vector and differentiation vector with this metric.

4. (iv)

   PS calculation with randomized GRN model to calculate PS cut-off value

CellOracle also produces randomized GRN models. The randomized GRNs can be used to generate dummy negative control data in CellOracle simulations. We calculated cut-off values for the negative PS analysis in the systematic KO simulation. First, the negative PS is calculated for all TFs using either a normal or a randomized vector. The score distribution generated from the randomized vector was used as a null distribution. We determined the cut-off value corresponding to a false-positive rate of 0.01 by selecting the 99th percentile value of PSs generated with randomized results (Extended Data Fig. 3g).

Network analysis

In addition to CellOracle's unique gene perturbation simulation, CellOracle's GRN model can be analysed with general network structure analysis methods or graph theory approaches. Before this network structure analysis, we filter out weak or insignificant connections. GRN edges are initially filtered on the basis of *P* values and absolute values of edge strength. The user can define a custom value for the thresholding according

to the data type, data quality and aim of the analysis. After filtering, CellOracle calculates several network scores: degree centrality, betweenness centrality and eigenvector centrality. It also assesses network module information and analyses network cartography. For these processes, CellOracle uses igraph (https://igraph.org).

Validation and benchmarking of CellOracle GRN inference

To test whether CellOracle can correctly identify cell-type- or cell-state-specific GRN configurations, we benchmarked our new method against diverse GRN inference algorithms: WGCNA, DCOL, GENIE3 and SCENIC. WGCNA is a correlation-based GRN inference algorithm, which is typically used to generate a non-directional network[53]; DCOL is a ranking-based non-linear network modelling method[54]; and GENIE3 uses an ensemble of tree-based regression models, and aims to detect directional network edges. GENIE3 emerged as one of the best-performing algorithms in a previous benchmarking study[55]. The SCENIC algorithm integrates a tree-based GRN inference algorithm with information on TF binding[7].

Preparation of input data for GRN inference

We used the Tabula Muris scRNA-seq dataset for GRN construction input data[56]. Cells were subsampled for each tissue on the basis of the original tissue-type annotation: spleen, lung, muscle, liver and kidney. Data for each tissue were processed using the standard Seurat workflow, including data normalization, log transformation, finding variable features, scaling, principal component analysis (PCA) and Louvain clustering. The data were downsampled to 2,000 cells and 10,000 genes using highly variable genes detected by the corresponding Seurat function. Cell and gene downsampling were necessary to run the GRN inference algorithms within a practical time frame: we found that some GRN inference algorithms, especially GENIE3, take a long time with a large scRNA-seq dataset, and GENIE3 could not complete the GRN inference calculation even after several days if the whole dataset was used.

GRN inference method

After preprocessing, the exact same data were subjected to each GRN inference algorithm to compare results fairly. We followed the package tutorial and used the default hyperparameters unless specified otherwise. Details are as follows. WGCNA: we used WGCNA v.1.68 with R 3.6.3. WGCNA requires the user to select a 'power parameter' for GRN construction. We first calculate soft-thresholding power using the 'pickSoftThreshold' function with networkType="signed". Other hyperparameters were set to default values. Using the soft-thresholding power value, the 'adjacency' function was used to calculate the GRN adjacency matrix. The adjacency matrix was converted into a linklist object by the 'getLinkLis' function and used as the inferred value of the WGCNA algorithm. DCOL: we used nlnet v.1.4 with R 3.6.3. The 'nlnet' function was used with default parameters to make the DCOL network. The edge list was extracted using the 'as_edgelist' function. DCOL infers an undirected graph without edge weights. We assigned the value 1.0 for the inferred network edge and 0.0 for other edges. The assigned value was used as the output of the DCOL algorithm. GENIE3: we used GENIE3 v.1.8.0 with R 3.6.3. The GRN weight matrix was calculated with the processed scRNA-seq data using the 'GENIE3' function and converted into a GRN edge and weight list by the 'getLinkList' function. GENIE3 provides a directed network with network weight. The weight value was directly used as the inferred value of the GENIE3 algorithm. SCENIC: we used SCENIC v.1.2.2 with R 3.6.3. The SCENIC GRN calculation involves multiple processes. The calculation was performed according to SCENIC's tutorial (https://rdrr.io/github/aertslab/SCENIC/f/vignettes/SCENIC_Running.Rmd). First, we created the initialize settings configuration object with 'initializeScenic'. Then we calculated the co-expression network using the 'runGenie3' function, following the GRN calculation with several SCENIC functions; runSCENIC_1_coexNetwork2modules, runSCENIC_2_createRegulons and runSCENIC_2_createRegulons. We used the '10kb' dataset for the promoter information range. The calculated GRN information was loaded with the 'loadInt' function, and the 'CoexWeight' value was used as the inferred value of the SCENIC algorithm.

Ground-truth data preparation for GRN benchmarking

Cell-type-specific ground-truth GRNs were generated in the same manner as in a previous benchmarking study[55]. Here, we selected tissues commonly available in the Tabula Muris scRNA-seq dataset, mouse sci-ATAC-seq atlas data and ground-truth datasets: heart, kidney, liver, lung and spleen. The ground-truth data were constructed as follows. (i) Download all mouse TF ChIP–seq data as bed files from the ChIP-Atlas database (https://chip-atlas.org). (ii) Remove datasets generated under non-physiological conditions. For example, we removed ChIP–seq data from gene KOs or adeno-associated virus treatment. (iii) Remove data that include fewer than 50 peaks. (iv) Select peaks detected in multiple studies. (v) Group data by TF and remove TFs if the number of detected target genes is less than ten peaks. (vi) Convert data into a binary network: each network edge is labelled either 0 or 1, representing its ChIP–seq binding between genes. These steps yielded tissue- or cell-type-specific ground-truth data for 80 TFs, corresponding to 1,298 experimental datasets.

GRN benchmarking results

GRN inference performance was evaluated by the AUROC and the early precision ratio (EPR), following the evaluation method used in a previous benchmarking study[55]. CellOracle and SCENIC outperformed WGCNA, DCOL and GENIE3 based on AUROC (Extended Data Fig. 2a). This is because CellOracle and SCENIC filter out non-transcriptional connections (that is, non-TF–target gene connections) and other methodologies detect many false-positive edges between non-TFs. CellOracle with a scATAC-seq atlas base GRN performed better than CellOracle with a promoter base GRN and SCENIC. This difference was mainly derived from sensitivity (or true-positive rate). With scATAC-seq data, CellOracle captures a higher number of regulatory candidate genes. Considering EPR, representing inference accuracy for top k network edges (k = number of network edges with the label '1' in the ground-truth data), CellOracle performed well compared to other approaches (Extended Data Fig. 2b): GENIE3 and WGCNA assigned a high network edge weight to many non-transcriptional connections, resulting in many false-positive edges for the highly ranked inferred genes.

The CellOracle GRN construction method was analysed further to assess the contribution of the base GRN. We performed the same GRN benchmarking with a scrambled motif base GRN or no base GRN. For the scrambled motif base GRN, we used scrambled TF-binding-motif data for the base GRN construction. For the no base GRN analysis, selection of regulatory candidate genes was skipped, and all genes were used as regulatory candidate genes. As expected, the AUROC scores decreased when we used the scrambled motif base GRN (ranked 12/13 in AUROC, 11/13 in EPR; Extended Data Fig. 2a,b), decreasing even further in the no base GRN model (13/13; Extended Data Fig. 2a,b). The scrambled motif base GRN did not detect many regulatory candidate TFs, producing lower sensitivity. However, the scrambled motif base GRN can still work positively by removing connections from non-TF genes to TFs, functioning to filter out false-positive edges, and resulting in a better score relative to the no base GRN model. In summary, the base GRN is primarily important to achieve acceptable specificity, and the scATAC-seq base GRN increases sensitivity.

Next, we used CellOracle after downsampling cells to test how cell number affects GRN inference results. Cells were downsampled to 400, 200, 100, 50, 25 and 10 cells and used for GRN analysis with the scATAC-seq base GRN. GRNs generated with 400, 200, 100 and 50 cells received comparable or slightly reduced AUROC scores. The AUROC score decreased drastically for GRNs generated with 25 and 10 cells (Extended Data Fig. 2c). EPR was relatively robust even with small cell numbers (Extended Data Fig. 2d).

We performed additional benchmarking to investigate data compatibility between the base GRN and scRNA-seq data sources. A tissue-specific base GRN was generated separately using bulk ATAC-seq data[57]. We focused on the same five tissue types as above. Unprocessed bulk ATAC-seq data were downloaded from the NCBI database using the SRA tool kit (spleen: SRR8119827; liver: SRR8119839; heart: SRR8119835; lung: SRR8119864; and kidney: SRR8119833). After FASTQC quality check (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), fastq files were mapped to the mm9 reference genome and converted into bam files. Peak calling using HOMER was used to generate bed files from the bam

files. Peak bed files were then annotated with HOMER. Peaks within 10 kb around the TSS were used. Peaks were sorted by the 'findPeaks Score' generated by the HOMER peak-calling step, and we used the top 15,000 peaks for base GRN construction. These peaks were scanned with the gimmemotifs v.5 vertebrate motif dataset, which is the same motif set we use for scATAC-seq base GRN construction.

We compared benchmarking scores between GRN inference results generated from different base GRNs. Overall, GRN construction performed best when the same tissue type for ATAC-seq base GRN construction and scRNA-seq was used (10/13 in AUROC, 11/13 in EPR; Extended Data Fig. 2e,f). The score was lower with different tissue types combined between the base GRN and scRNA-seq data. In summary, benchmarking confirmed that our GRN construction method performs well for the task of transcriptional GRN inference.

CellOracle evaluation

Evaluation of simulation value distribution range

We investigated a range of simulated values to confirm that the signal propagation step does not generate an out-of-distribution prediction. Specifically, we assessed the distribution of the sum of the simulated shift and original gene expression, which correspond to the simulated expression level (termed 'simulation gene expression level' here for explanatory purposes: $X_{\text{simulation gene expression level}} = X_{\text{original}} + \Delta X_{\text{simulated,}}$). We evaluate all genes, comparing the simulation gene expression level with the original gene expression distribution. To detect out-of-distribution data, we calculated the maximum exceedance percentage, representing the percentage difference of the maximum value of the simulated gene expression level compared to the maximum value of the wild-type gene expression value. The higher maximum exceedance indicates a bigger difference between simulated and wild-type values, identifying out-of-distribution values. For the *Spi1* KO simulation with the Paul et al. haematopoiesis dataset[16], we present the top four genes showing the maximum exceedance values (Supplementary Fig. 2). The simulation

expression levels of even these genes appear very similar to the original wild-type distributions of gene expression. For example, in the *Ly86* simulated value distribution, 99.963% of all cells are within the wild-type gene expression range. Only 0.037% of cells exhibit a *Ly86* gene simulation value outside the wild-type distribution, but the maximum difference is only 3.2%. We designed CellOracle to simulate a minimal relative shift vector rather than an out-of-distribution prediction, confirmed by this analysis. The functions we have used for these analyses are implemented in CellOracle. Users can check simulation value distributions, and CellOracle will produce a warning if out-of-distribution simulations occur.

To further explore the minimum number of cells with minor out-of-distribution values, we generated a simulation vector in which the out-of-distribution values are clipped into the wild-type distribution range. The simulated cell-identity shift vector of clipped values is indistinguishable compared to the original results (Supplementary Fig. 2b–e), confirming that the CellOracle simulation is not relying on these out-of-distribution values. The out-of-distribution value can be clipped if we add 'clip_delta_X=True' in the CellOracle signal propagation function. Thus, users can ensure the simulation is not relying on out-of-distribution values.

CellOracle simulation results generated with randomized GRN or no signal propagation

We performed KO simulation with randomized GRN models to clarify the necessity of the GRN signal propagation simulation. In addition, we calculated cell-identity vectors without the signal propagation step; the cell-identity shift vector was calculated solely on the basis of input TF expression loss, thus representing the information from the expression pattern of only a single TF. The vector map in Supplementary Fig. 3 shows *Gata1* KO simulation results and *Spi1* KO simulation results with an intact GRN coefficient matrix, randomized GRN matrix or no GRN signal propagation. The randomized GRN analysis results and no GRN signal propagation results show only slight cell-identity shift vectors (Supplementary Fig. 3b,c,e,f). Although very subtle vectors can be observed, most expected simulation results are not obtained. Thus, we

confirmed that the GRN signal propagation strategy has an essential role in the CellOracle KO simulation.

Evaluation of signal propagation number

We next tested the number of iterations at the signal propagation step. We performed KO simulations using two independent mouse haematopoiesis datasets: Paul et al.[16] and Dahlin et al.[58]. For several TFs, we tested different numbers of signal propagation rounds in the KO simulations across independent clusters. First, focusing on the Paul dataset, simulation vector fields for *Spi1* and *Gata1*, with 0, 1 and 3 rounds of signal propagation, were investigated (Supplementary Fig. 4). The simulation under hyperparameter $n = 0$ shows the vector calculated without any signal propagation within the GRN; that is, the vector is calculated from only the difference of the input TF gene expression shift. This $n = 0$ simulation shows almost no phenotype, showing the necessity of the GRN signal propagation process. Next, a comparison of vector fields from $n = 1$ and $n = 3$ simulations shows similar results. This is not surprising given the following. (1) Most coefficient values in the GRN are small, ranging between −1 and 1 (Supplementary Fig. 4d). (2) Accordingly, the signal will be attenuated over the propagation process in most cases. (3) This also means that the first signal propagation step will produce the most significant shifts relative to the later steps. However, when scrutinizing the vectors, we observe a more evident shift in cell identity around the late GMP cluster and the early granulocytes in the $n = 3$ *Gata1* KO vectors compared to $n = 1$ vectors. The results suggest that the second and third rounds of signal propagation increase the sensitivity to detect small shifts by adding the second and third rounds of downstream gene effects, respectively.

To quantify these observations and determine whether there is an ideal number of signal propagation rounds, we investigated the L1-norm of ΔX, representing the sum of the magnitudes of each simulated gene expression shift. The L1-norm of ΔX is almost saturated at the $n = 3$ in most cases (Supplementary Fig. 4c). We also performed these analyses with the Dahlin haematopoiesis dataset[58] (Supplementary Fig. 5). Overall, the results are consistent with our analysis of the Paul data. Again, we observe that the L1-

norm of ΔX is saturated at relatively small n values in most cases. However, *Cebpa* is an outlier in this analysis, in which the delta X length gradually and continuously increases as n increases. We next examined the vector field of *Cebpa* with various n (Supplementary Fig. 6). Despite such divergence of the L1-norm of ΔX, the vector field of *Cebpa* showed consistent results, suggesting that the calculation strategy for cell-identity shift is robust using the local neighbour vectors (Extended Data Fig. 1e).

Altogether, at $n = 3$, the simulated shift vectors almost converge, producing consistent results. In rare cases, the L1-norm of ΔX might show divergence with n. However, the $n = 3$ simulation results are consistent with higher n values, which might generate unexpected behaviour owing to signal divergence. On the basis of these analyses, we recommend that users perform three iterations for the signal propagation step.

Selection of dimensionality reduction method

CellOracle simulation with UMAP and *t*-SNE using Paul et al. haematopoiesis data

We used UMAP and *t*-distributed stochastic neighbour embedding (*t*-SNE) for the perturbation simulation analysis to show how the choice of dimensionality reduction affects CellOracle results. We used Scanpy to construct UMAP or *t*-SNE plots using the Paul et al. haematopoiesis dataset[16]. In the UMAP (Supplementary Fig. 7a), we observe a similar trajectory that agrees with the force-directed graph (Fig. 1b). However, monocyte and granulocyte branches on the UMAP are relatively less resolved. This notwithstanding, the simulation results using the UMAP (Supplementary Fig. 8, top) lead to the same conclusion as Fig. 1. For example, in the *Gata1* KO simulation, we correctly predict inhibited differentiation along the MEP lineage whereas GMP differentiation is promoted. Furthermore, we predict inhibited GMP to granulocyte differentiation, consistent with our force-atlas-based presentation in Fig. 1h. In comparison, the overall structure of the *t*-SNE graph is consistent with the force-directed and UMAP graphs, although it lacks resolution (Supplementary Fig. 7b). However, the *t*-SNE results still agree with Fig. 1, just at a lower resolution (Supplementary Fig. 8, bottom). In conclusion, we

stress that the choice of the dimensional reduction algorithm is crucial to sensitively analyse the cell differentiation trajectory.

Guidance for selecting the dimensionality reduction method

For the force-directed graph calculation, we recommend using Scanpy's sc.pl.draw_graph function[59] or SPRING[60]. Both internally use force atlas 2 (ref. [61]). Compared to UMAP, force-directed graphs can capture more fine-branching structures but can be unstable if the data have many branches that can overlap. To avoid branch overlap, PAGA cell trajectory information can be used to initiate the force-directed graph calculation: https://scanpy.readthedocs.io/en/stable/tutorials.html#https://github.com/theislab/paga.

We recommend using force-directed graphs as a first choice because they generally produce a high-resolution lineage structure. However, we recommend UMAP as a reliable alternative if overlapping branches are observed. In our CellOracle tutorial, we show the detailed guide and code for the dimensionality reduction implementation, including data preprocessing: https://morris-lab.github.io/CellOracle.documentation.

CellOracle KO simulation with unrelated cell-type base GRNs

To assess how base GRN performance relates to scATAC data source, we performed TF KO simulations in haematopoiesis using the 'general' mouse scATAC-seq atlas[13] base GRN versus a heart-specific base GRN to represent an unrelated cell type (Supplementary Fig. 9). The simulation vectors using the mismatched heart base GRN are weaker, although still in general agreement. We reason that even if the base GRN retains some edges that are not active in the scRNA-seq data, CellOracle can still work robustly. However, simulation with the heart base GRN fails to detect the early granulocyte phenotype in the *Gata1* KO and almost all shifts in the *Cepba* KO, suggesting reduced sensitivity with the mismatched base GRN.

We also assess the mean degree centrality (the number of genes to which a TF is connected) in the inferred GRNs for each of four TFs (Supplementary Fig. 10). With the inappropriate heart base GRN, CellOracle fails to build

network edges for some genes, resulting in a low degree centrality score and reduced simulation sensitivity. We recommend starting CellOracle analysis with the general GRN and comparing its performance against tailored base GRNs.

Markov simulation based on CellOracle simulation vector

To estimate cell distribution in response to gene perturbation, we need to consider both the differentiation hierarchy and the perturbation vector together. We performed a Markov random walk simulation as described previously[52] (https://github.com/velocyto-team/velocyto.py) with some modifications. First, our Markov simulation used the CellOracle cell-identity transition vector in addition to the differentiation vector; the transition probability matrix for these vectors was applied alternatively to consider both effects. Second, cells in early differentiation stages were selected and used for the initial state of our Markov simulation, whereas the previous study used the whole population as the initial state[52]. The Markov simulation analysis with data from another study[59] is shown in Supplementary Fig. 17 to show typical simulation results and their interpretation.

CellOracle analysis with previously published scRNA-seq and scATAC-seq data

Paul et al. mouse haematopoiesis scRNA-seq data

The GEM was downloaded with Scanpy's data loading function, scanpy.datasets.paul15(). After removing genes with zero counts, the GEM was normalized by total UMI counts ((scanpy.pp.filter_genes(min_counts=1), scanpy.pp.normalize_per_cell(key_n_counts='n_counts_all')). Highly variable genes, including 90 TFs, detected by scanpy.pp.filter_genes_dispersion(flavor='cell_ranger', n_top_genes=2000, log=False), were used for the following downstream analysis: the GEM was log-transformed, scaled and subjected to PCA (scanpy.pp.log1p(), scanpy.pp.scale(), scanpy.tl.pca(svd_solver='arpack')). We calculated the

force-directed graph dimensional reduction data based on the PAGA graph[62] for initialization (scanpy.tl.paga(), scanpy.tl.draw_graph(init_pos='paga')). Cells were clustered using the Louvain clustering method (scanpy.tl.louvain (resolution=1.0)). Clusters were annotated manually using marker gene expression and the previous annotations from Paul et al.[16] We removed dendritic cell (DC) and lymphoid cell clusters to focus on myeloid cell differentiation. GRN calculation and simulation were performed as described above, using the default parameters. For the base GRN, we used the base GRN generated from the mouse sci-ATAC-seq atlas dataset[13].

Cell density was visualized using a kernel density estimation (KDE) plot. First, we performed random downsampling to 768 cells to adjust the cell number between WT and KO samples. KDE was calculated with the scipy.stat.gaussian_kde function. The calculated KDE was visualized with the matplotlib.pyplot.contour function. We used the same contour threshold levels between all samples.

Although we did not focus on the network structure in the main text, we examined CellOracle GRN models using graph theory approaches before the simulation analysis. Graph theory analysis revealed that these inferred GRN configurations resemble a scale-free network the degree distribution of which follows a power law, a characteristic configuration of biological networks[63] (Extended Data Fig. 3b). Further, we assess GRNs using degree centrality—a basic measure of how many genes a TF connects to[63]. Using the MEP cluster as an example, 27 out of 30 genes with a high degree centrality score in the MEP_0 GRN are confirmed known regulators of MEP lineage differentiation or stem and progenitor cell function (Extended Data Fig. 3c and Supplementary Table 2). Analysis of additional clusters yielded similar agreement with previous literature, confirming that CellOracle GRN inference captures biologically plausible cell-state-specific GRN structures, consistent with previous biological knowledge. All network analysis and simulation results can be explored at https://www.celloracle.org.

Pijuan-Sala et al. mouse early gastrulation and organogenesis scRNA-seq data

We applied CellOracle to a scRNA-seq atlas of mouse gastrulation and organogenesis by Pijuan-Sala et al.[30]. This single-cell profiling of WT cells highlighted a continuous differentiation trajectory across the early development of various cell types (Extended Data Fig. 9a). In addition, the developmental effects of *Tal1* KO, a TF known to regulate early haematoendothelial development[64,65], were investigated in this study. We validated the CellOracle simulation using these *Tal1* KO ground-truth scRNA-seq data. The data were generated from seven chimeric E8.5 embryos of WT and *Tal1* KO cells (25,307 cells and 26,311 cells, respectively). We used the R library, MouseGastrulationData (https://github.com/MarioniLab/MouseGastrulationData), to download the mouse early gastrulation scRNA-seq dataset. This library provides the GEM and metadata. We used the *Tal1* chimera GEM and cell-type annotation, "cell type.mapped", provided by this library. Data were normalized with SCTransform[66]. The GEM was converted to the AnnData format and processed in the same way as the Paul et al. dataset. For the dimensionality reduction, we used UMAP using the PAGA graph for the initialization (maxiter=500, min_dist=0.6). We removed the extraembryonic ectoderm (ExE), primordial germ cell (PGC) and stripped nuclei clusters which lie outside the main differentiation branch. After removing these clusters, we used the WT cell data for the simulations (24,964 cells). GRN calculations and simulations were performed as described above using the default parameters. We used the base GRN generated from the mouse sci-ATAC-seq atlas dataset. We constructed cluster-wise GRN models for 25 cell states. Then, we simulated *Tal1* KO effects using the WT scRNA-seq dataset. For the late-stage-specific *Tal1* conditional KO simulation, we set *Tal1* expression to be zero in the blood progenitor and erythroid clusters to analyse the role of *Tal1* in late erythroid differentiation stages (Extended Data Fig. 9i,j).

Farrell et al. zebrafish early development scRNA-seq data

GEM, metadata and URD trajectory data were downloaded from the Broad Institute Single Cell Portal (https://tinyurl.com/7dup3b5k). We used the cell clustering data and developmental lineage data from Farrell et al.[32] The GEM was already normalized and $\log_2$-transformed, which we converted to

non-log-transformed data before CellOracle analysis. The GEM had human gene symbols, which we converted back to zebrafish gene symbols using gene name data in ZFIN (https://zfin.org). We used URD dimensional reduction embedding data. To use the URD differentiation trajectory in the CellOracle simulations, we ran several preprocessing and calculations. We first identified cells with URD coordinate data ($n = 26{,}434$ cells). The "EPL/periderm and primordial germ cell" cluster, which represents 1.7% of the total population, was also excluded from our analysis because it is located outside the main differentiation trajectory branch. The whole URD structure ($n = 25{,}711$ cells) was split into four sub-branches to simplify the calculations (Extended Data Fig. 10b). Then, we converted the original URD coordinates, a 3D matrix, into a 2D matrix using PCA (sklearn.decomposition.PCA) because CellOracle requires 2D dimensional reduction embedding data. The GEM was converted into the AnnData format. At the variable gene detection step, we selected the top 3,000 genes. GRN calculation and simulations were performed as described above using the default parameters. We did not calculate pseudotime because the pseudotime data calculated with URD were available. The pre-calculated pseudotime data were used to calculate the 2D development vector field. For base GRN construction, we used UCSC TSS and promoter data and the zebrafish reference genome (https://useast.ensembl.org/Danio_rerio/Info/Index), danRer11 (the bed file is included in the CellOracle package). The promoter DNA sequence was scanned with CisBP version2 motif dataset to generate the base GRN (http://cisbp.ccbr.utoronto.ca).

For screening novel regulators of axial mesoderm cell identity, we prioritized candidate genes as follows. First, we performed CellOracle KO simulations for 232 active TFs, which had at least one gene edge in the constructed GRN model in the axial mesoderm branch (Extended Data Fig. 12a, step 1). We focused on the early differentiation stage by selecting cells between digitized pseudotime 0 and 3 (Extended Data Fig. 12a, step 2). For this analysis, we focused on negative perturbation scores to identify candidate TFs. A large negative perturbation score indicates a predicted inhibition or block in differentiation following TF KO; thus, we reasoned that these TFs might have a positive role in differentiation (Extended Data Fig. 12a, step 3). To prioritize TFs according to the predicted differentiation

inhibition effects, we ranked TFs according to the sum of their negative perturbation scores, resulting in the 30 genes listed in Fig. 5a. Next, we surveyed the GRN degree centrality scores of 30 candidate genes in the notochord cluster GRN because we reasoned that those genes with higher GRN degree centrality result in a more reliable simulation. Then, we calculated the gene specificity score comparing the axial mesoderm sub-branch and the other sub-branches using the Scanpy function, sc.tl.rank_genes_groups(). Although gene specificity does not necessarily relate to gene function, we assumed that specific gene expression would simplify the interpretation of experimental results and reduce the likelihood of unexpected phenotypes from clusters other than axial mesoderm. Finally, we analysed mean expression, assuming perturbation experiments with highly expressed genes would be more robust, especially in the CRISPR–Cas9-based F_0 embryo analysis. After removing previously reported genes, we selected candidate genes that exist in the 50th percentile of these scores (Extended Data Fig. 12b, highlighted in a grey rectangle), resulting in *lhx1a*, *sebox*, *irx3a*, *creb3l1 and zic2a*. We finally selected three candidates, *lhx1a*, *sebox* and *irx3a*, and removed *creb3l1* and *zic2a* from the first LOF experiment list, according to the following rationale: *creb3l1* gene sequences are similar to *creb3l2*; thus, it was challenging to design specific sgRNAs to target this gene; *creb3l2* was previously reported to regulate axial mesoderm development. Although *zic2a* narrowly passed the gene specificity threshold described above, we found that *zic2a* expression was high in the other mesendoderm sub-branch and the ectoderm sub-branches; thus, we excluded this gene from our downstream analyses.

Dahlin et al. mouse haematopoiesis scRNA-seq data

Mouse haematopoiesis scRNA-seq data from Dahlin et al.[58] were downloaded from the PAGA GitHub repository (https://github.com/theislab/paga). The GEM was normalized by total UMI counts after removing genes with zero counts ((scanpy.pp.filter_genes(min_counts=1), scanpy.pp.normalize_per_cell(key_n_counts=‘n_counts_all’)). Highly variable genes were detected and used for the following downstream analysis: (scanpy.pp.filter_genes_dispersion(flavor=‘cell_ranger’,

n_top_genes=3000, log=False)). The GEM was log-transformed, scaled and subjected to PCA and Louvain clustering (scanpy.pp.log1p(), scanpy.pp.scale(), scanpy.tl.pca(svd_solver='arpack'), scanpy.tl.louvain(resolution=1.5)). The original force-directed graph reported in Dahlin et al.[58] was used for the CellOracle simulation. GRN calculation and simulation were performed using the default parameters. For the base GRN, we used the mouse sci-ATAC-seq atlas dataset[13].

Setty et al. human haematopoiesis scRNA-seq data

Human haematopoiesis scRNA-seq were downloaded from the Human Cell Atlas: https://data.humancellatlas.org/explore/projects/091cf39b-01bc-42e5-9437-f419a66c8a45 (Setty et al.)[67]. The GEM was normalized by total UMI counts after removing genes with zero counts ((scanpy.pp.filter_genes(min_counts=1), scanpy.pp.normalize_per_cell(key_n_counts='n_counts_all')). Highly variable genes were detected and used for the following downstream analysis: (scanpy.pp.filter_genes_dispersion(flavor='cell_ranger', n_top_genes=3000, log=False)). The GEM was log-transformed, scaled and subjected to PCA and Louvain clustering (scanpy.pp.log1p(), scanpy.pp.scale(), scanpy.tl.pca(svd_solver='arpack'), scanpy.tl.louvain(resolution=1.5)). The force-directed graph was calculated with SPRING (https://kleintools.hms.harvard.edu/tools/spring.html). We removed DC and lymphoid cell clusters in line with the Paul et al.[16] data analysis. GRN calculation and simulation were performed using the default parameters. For the base GRN, we used the base GRN generated using the Buenrostro et al. human haematopoiesis scATAC-seq data described below[68].

Buenrostro et al. human haematopoiesis scATAC-seq data

Human haematopoiesis scATAC-seq data from Buenrostro et al.[68] were used to construct a human haematopoiesis base GRN. The scATAC-seq peak data and count matrix was obtained from the Gene Expression Omnibus (GEO), with accession code GSE96769, and processed with Cicero (v.1.3.4) to obtain co-accessibility scores as follows: After removing

peaks with zero counts, cells were filtered by the peak count (min count = 200, max count = 30,000). The data were processed using Cicero functions (detect_genes(), estimate_size_factors(), preprocess_cds(method = "LSI"), reduce_dimension(reduction_method = 'UMAP', preprocess_method = "LSI")). Then Cicero co-accessibility scores were calculated using run_cicero() with human chromosome length information imported by data("human.hg19.genome"). Output peak and co-accessibility scores were used for CellOracle base GRN construction. CellOracle annotated the TSS site in the peaks, and the TSS peaks and *cis*-regulatory peaks with co-accessibility scores ≥ 0.8 were used for motif scanning. We used the gimmemotifs vertebrate v5 motif dataset, which is CellOracle's default for mouse and human motif scanning.

TF motif enrichment analysis was performed using ChromVar[68]. The ChromVar score matrix, which includes 2,034 cells and 1,764 motif data, was processed with scanpy to generate a force-directed graph and Louvain clustering (scanpy.tl.pca(), scanpy.tl.louvain(resolution=0.5), scanpy.tl.draw_graph()). The cluster was annotated using cell source FACS gate sample labels. The score fold change was calculated and visualized as a volcano plot (Supplementary Fig. 16). The statistical test was performed using the two-tailed Wilcoxon rank-sum test with Bonferroni correction.

Comparison between CellOracle haematopoiesis KO simulation results and previous reports

CellOracle KO simulation results for 12 key TFs that regulate myeloid differentiation are shown in Figs. 1 and 2, Extended Data Figs. 5 and 6 and Supplementary Figs. 13 and 14. The simulation results were compared with previous reports (summarized in Supplementary Table 2). In these figures, the summary of the simulation results is shown in the right column with the mark (*), which indicates that the simulation results agree with the previously reported role or phenotype of the TF. We note that the input haematopoiesis data focus on the myeloid lineage; thus, the simulation results show relative cell-identity shifts within the myeloid lineage only. For example, *Spi1* has an important role not only in the myeloid lineage but also in other cell types, such as HSCs and lymphoid lineages[69]. However, we

cannot simulate the role in these cell types if they are not present in the input data.

1. (1)

   Klf1 (*KLF1*)

   Klf1 promotes differentiation towards the ME lineage, promoting erythroid cell differentiation in particular[15]. CellOracle simulation results agree with this role (Extended Data Fig. 5a and Supplementary Figs. 13e and 14e).

2. (2)

   Gata1 (*GATA1*)

   Gata1 promotes ME lineage differentiation and also promotes granulocyte differentiation[15,70]. Both the Paul et al.[16] and Dahlin et al.[58] data simulation results reproduce these *Gata1* roles. (Fig. 1f and Supplementary Fig. 13b). In the Setty et al. dataset[67], the ME lineage phenotype is reproduced, but the granulocyte phenotype is not observed (Supplementary Fig. 14b). We speculate that this is because the Setty dataset includes few mature granulocytes.

3. (3)

   Gata2 (*GATA2*)

   Gata2 is a key factor in maintaining stemness in MPPs[15]. Simulation results in all data agree with this role for *Gata2* (Extended Data Fig. 6a and Supplementary Figs. 13i and 14g).

4. (4)

   Spi1 (*SPI1*)

   Spi1 promotes GM lineage differentiation. The inhibition of *Spi1* shifts cell identity from the GM to the ME lineage[15,71]. Simulation results in

all datasets agree with this role of *Spi1* (Fig. 1e and Supplementary Figs. 13a and 14a).

5. (5)

   Cebpa (*CEBPA*)

   Cebp*a* promotes GM lineage differentiation while inhibiting ME lineage differentiation[16,72], and promoting granulocyte differentiation in particular[15]. Simulation results using the Paul et al.[16] and Dahlin et al.[58] datasets agree with this role for *Cebpa* (Fig. 2b and Supplementary Fig. 13c). Although the ME lineage phenotype is not detected using the Setty et al. dataset[67], the GM lineage phenotype is successfully reproduced (Supplementary Fig. 14c).

6. (6)

   Cebpe (*CEBPE*)

   Cebpe promotes granulocyte lineage differentiation[15,16]. Simulation results in all datasets agree with this role of *Cebpe* (Fig. 2c and Supplementary Figs. 13d and 14d).

7. (7)

   Gfi1 (*GFI1*)

   Gfi1 promotes granulocyte lineage differentiation[15,72,73,74]. Simulation results using the Paul et al.[16] and Dahlin et al.[58] datasets agree with this role of *Gfi1* (Extended Data Fig. 5c and Supplementary Fig. 13g).

8. (8)

   Gfi1b (*GFI1B*)

   Gfi1b promotes ME lineage differentiation[15]. Simulation results in all data suggest that *Gfi1b* promotes erythroid differentiation (Extended Data Fig. 5b and Supplementary Fig. 13f). The Mk phenotype is

unclear in the simulation, probably owing to the small numbers of Mk cells.

9. (9)

   Irf8 (*IRF8*)

   Irf8 promotes GM lineage differentiation. In particular, *Irf8* promotes monocyte differentiation as a lineage switch between monocyte and granulocyte bifurcation[29]. Simulation results in all data agree with the role of *Irf8* (Extended Data Fig. 5d and Supplementary Figs. 13h and 14f).

10. (10)

    Lmo2 (*LMO2*)

    Lmo2 is a central factor in maintaining stemness in the MPP compartment[15]. Simulation results using the Dahlin et al. data[58] agree with this role. (Supplementary Fig. 13l). However, simulation results using Paul et al. data[16] showed a different phenotype in erythrocyte cells, suggesting that *Lmo2* is also crucial for promoting erythroid differentiation (Extended Data Fig. 6d). A function of *Lmo2* in promoting erythroid differentiation was also reported[75].

11. (11)

    Runx1 (*RUNX1*)

    Runx1 is an central factor in maintaining stemness in the MPP compartment[15] Simulation results in all datasets agree with this role of *Runx1* (Extended Data Fig. 6b and Supplementary Fig. 13j).

12. (12)

    Fli1 (*FLI1*)

Fli1 has context-dependent roles. *Fli1* is a key factor for Mk differentiation[15], and for maintaining stemness in the stem and progenitor comparment[76]. The simulations consistently reproduce these phenotypes (Extended Data Fig. 6c and Supplementary Figs. 13k and 14h). In addition, a previous study reported that loss of *Fli1* causes dysregulation in later differentiation stages[77], consistent in simulations using the Paul et al. dataset[16] (Extended Data Fig. 6c).

Zebrafish lines

The zebrafish experiments were approved by the Institutional Animal Care and Use Committees at Washington University in St Louis. All animal experiments followed all relevant guidelines and regulation. The following zebrafish lines were used in this study: AB* and *floating head*$^{n1/n1}$ (*flh/noto*) mutants[37]. Sample sizes and developmental stages are stated below. Randomization was not performed as experimental groups were determined by genotype. Blinding was performed for the generation and analysis of the single-cell data.

CRISPR–Cas9-based mutagenesis of F_0 embryos

To generate somatic gene deletions in early zebrafish embryos, we used CRISPR–Cas9 with two or three sgRNAs as described previously[78]. In brief, sgRNAs were designed using CHOPCHOP (http://chopchop.cbu.uib.no/) to target 5′ exons and the functional domain of each selected TF and synthesized (IDT) (Supplementary Fig. 20b). sgRNA sequences are listed in Supplementary Table 6. Duplex sgRNA was prepared by mixing equimolar amounts of Alt-R crRNA and Alt-tracrRNA (IDT) in IDT Duplex Buffer, heating to 95 °C and slowly cooling to room temperature (RT) for 20 min. For the final mix of ribonucleoprotein complex (RNPs), around 4 µM duplex sgRNA was assembled with around 5 µM CRISPR–Cas nuclease (Alt-R S.p. HiFi Cas9 Nuclease V3) in 3 M KCl 0.025% and phenol red solution. The activity of HiFi Cas9 and selected sgRNAs was confirmed with regular PCR, Sanger sequencing and capillary electrophoresis, as described previously[40]. In brief, DNA from eight embryos for each combination of Cas9 and sgRNAs was extracted at

10 hpf. PCR amplification was performed with primers complementary to sequences 250 bp upstream and downstream of the PAM sequences (Supplementary Table 6). In addition, tracking of indels by decomposition (TIDE)[79] analysis was used to predict the percentage of indels at the target locus (Supplementary Fig. 20c). *flh*$^{n1/n1}$ mutant embryos were generated by crossing heterozygotes and selecting mutants on the basis of their morphology at 10 hpf.

Embryo collection and processing

Zebrafish embryos were produced by natural matings and injected at the one-cell stage with around 2–4 nl of RNP solution into the blastodisc. Embryos were incubated at 28 °C after removing those damaged during the injection process. After 9 hpf, embryos were enzymatically dechorionated and deyolked as previously described[32]. In brief, embryos were dechorionated by incubation in 1 mg ml^{-1} pronase, washed with 'blue water' and then transferred into plastic Petri dishes coated with 2% agarose with methylene blue water. Deyolking was performed manually by 'squeezing' the yolk out of the blastoderm cap with a closed pair of forceps inserted between the embryonic blastoderm and the yolk. The layer of cells detached from the yolk was transferred to a 1.5-ml Eppendorf tube with 50 µl of DMEM/F12 medium. For each experiment, 40–50 individual CRISPR–Cas9-targeted embryos (crispants) were prepared for dissociation into single-cell suspensions. Cell dissociation was performed according to the previous report (Farrell et al.)[32]. DMEM/F12 medium was added to the Eppendorf tube to bring the total volume to 200 µl. Cells were mechanically dissociated by flicking the tube 15 times and pipetting 3 times. The cell mixture was spun at 300*g* for 2 min and twice washed with PBS + 0.1% BSA. The same procedure was followed to collect and dissociate cells from WT and *flh*$^{n1/n1}$ mutant embryos.

RNA extraction and qRT-PCR

Total RNA was extracted from approximately 50 embryos for each experimental condition, homogenized in Trizol (Life Technologies) and further purified following Qiagen RNEasy Mini Kit instructions[80]. One microgram of total RNA was used to synthesize cDNA with the iScript kit (BioRad) following the manufacturer's protocol. SYBR green (BioRad) qRT-PCR reactions were run in a CFX Connect Real-Time PCR detection system (BioRad) with three technical replicates. The primers used are listed in Supplementary Table 6.

Whole-mount in situ hybridization

An antisense RNA probe for *nog1* was generated from plasmid pBSKII[81], previously linearized with Not1, and used as a template for in vitro transcription using NEB T7 RNA polymerase and RNTPs labelled with digoxygenin (DIG) (Roche). WISH was performed according to a previous report[82]. In brief, embryos were fixed overnight in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS), rinsed in PBS + 0.1% Tween 20 (PBT) and dehydrated in methanol. Embryos were then rehydrated in PBT and incubated for at least 2 h in hybridization solution (HM) with 50% formamide (in 0.75 M sodium chloride, 75 mM sodium citrate, 0.1% Tween 20, 50 μg ml^{-1} heparin (Sigma) and 200 μg ml^{-1} tRNA) at 70 °C, then hybridized overnight at 70 °C with antisense probes diluted approximately 1 ng μl^{-1} in hybridization solution. Embryos were washed through a series of 10 min, 70 °C washes in HM diluted with 2× SSC buffer (0.3 M sodium chloride and 30 mM sodium citrate) once in each of the following: 75% HM, 50% HM, 25% HM and 100% 2× SSC. The same gradual change from SSC to PBT was performed for the subsequent washes. Embryos were blocked at RT for several hours in PBT with 2% goat serum and 2 mg ml^{-1} bovine serum albumin (BSA), then incubated overnight at 4 °C with anti-DIG antibody (Roche 11093274910) at 1:5,000 on a horizontal shaker (40 rpm). Embryos were rinsed six times for 15 min per wash in PBT, and then in staining buffer (PBT+100 mM Tris pH 9.5, 50 mM $MgCl_2$ and 100 mM NaCl) before staining with BM Purple solution (Roche).

HCR

HCR was performed according to the protocols provided by Molecular Instruments (https://www.molecularinstruments.com). Embryos were fixed at 10 hpf with 4% PFA, dehydrated with methanol and rehydrated as described for WISH above. Embryos were pre-hybridized in hybridization buffer (Molecular Instruments) for 1 h at 37 °C and subsequently incubated in 200 µl of hybridization solution containing 1 pg of probes overnight at 37 °C. Embryos were then washed four times in wash buffer (Molecular Instruments) followed by two washes in 5× SSCT, containing 5× SSC buffer (Thermo Fisher Scientific) and 0.1% Tween 20 (Sigma). For the pre-amplification step, embryos were incubated in amplification buffer (Molecular Instruments) for more than 1 h. At the same time, hairpin mixtures were prepared by heating 12 pmol of hairpin 1 (H1) and 2 (H2) for each sample to 95 °C for 90 s, followed by cooling in the dark for 30 min at RT. H1 and H2 were mixed and then added to 200 µl amplification buffer. Embryos were incubated in the hairpin mixture at RT overnight in the dark. On the third day, embryos were washed more than 4 times in 5× SSCT and either stored at 4 °C or mounted for microscopy.

Microscopy

Embryos subjected to HCR were mounted in 3% low-melt agarose in glass-bottomed 35-mm Petri dishes. Alternatively, embryos were manually deyolked and flattened on a glass slide with one to two drops of 3% methylcellulose. Images of the anterior and posterior body regions were taken by acquiring around 200-µm *z*-stacks with a 1-µm step, using a 10× objective lens on a modified Olympus IX81 inverted spinning disc confocal microscope equipped with Voltran and Cobolt steady-state lasers and a Hamamatsu ImagEM EM CCD digital camera.

Image quantification with IMARIS software

Individual confocal 3D datasets were analysed with IMARIS 9.9 software (Bitplane). On the basis of the DAPI signal, radii were determined by taking half of the longest diameter of each nucleus, which was measured as a single spot using the ‘spots’ view in IMARIS. These parameters were

applied in all images used for quantification. Nuclei positive for specific probes within a selected area were identified using the 'spots' view as spots with a signal in the specific channel that overlapped with DAPI spots. Analysis was performed on eight embryos: four anterior and four posterior per experimental group.

10X Chromium procedure

For single-cell library preparation on the 10X Genomics platform, we used: the Chromium Single Cell 3′ Library & Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3′ Chip kit v2 (PN-120236) and Chromium i7 Multiplex Kit (PN-120262), according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagents Kits V2 User Guide. Before cell capture, methanol-fixed cells were placed on ice, then spun at 3,000 rpm for 5 min at 4 °C, followed by resuspension and rehydration in PBS, as described previously[83]. A total of 17,000 cells were loaded per lane of the chip, aiming to capture 10,000 single-cell transcriptomes. The resulting cDNA libraries were quantified on an Agilent TapeStation and sequenced on an Illumina NextSeq 550.

10X Chromium scRNA-seq data processing

10X alignment and digital GEM generation

The Cell Ranger v5.0.1 pipeline (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) was used to process data generated using the 10X Chromium platform. Cell Ranger processes, filters and aligns reads generated with the Chromium single-cell RNA sequencing platform. Following this step, the default Cell Ranger pipeline was implemented, and the filtered output data were used for downstream analyses.

Zebrafish scRNA-seq data processing

We used the R package Seurat (v.4.0.1) to process scRNA-seq data. Cells were filtered by RNA count and percentage of mitochondrial genes to

remove low-quality cells. Data were normalized using the Seurat NormalizeData() function. Variable genes were identified using the FindVariableFeatures() function with nfeature = 2,000. Data were integrated by applying Seurat functions, SelectIntegrationFeatures(), FindIntegrationAnchors() and IntegrateData() using default parameters. After data scaling, PCA and clustering were performed. The data after cell calling may include cells with very low mRNA counts generated from non-cell GEMs and ambient RNA. To remove such non-cell GEM data, we assessed the RNA count distribution to remove clusters with an abnormal RNA count distribution. Scaling, PCA, clustering and *t*-SNE were performed again after removing the cells above. *t*-SNE was calculated using the first 30 principal components. We applied the same pipeline to the WT reference, *flh* mutant and crispant scRNA-seq data.

After data integration and standard scRNA-seq preprocessing, the whole WT reference scRNA-seq data were annotated as follows. The segmentation labels generated in the Farrell et al.[32] zebrafish scRNA-seq data were transferred to the new scRNA-seq data using the Seurat function, FindTransferAnchors and TransferData, with default parameters. We manually adjusted the cell annotation to account for differences in the timing of cell collection. We generated cell-type annotations for the clustering data generated in the previous step by referring to the Farrell et al. dataset annotation labels. The WT reference cell-type annotations were transferred to the other scRNA-seq data using the same Seurat label transfer functions.

To compare cell identity on the same 2D embedding space, we used UMAP and the UMAP transfer function. We first calculated UMAP with axial mesoderm clusters in WT reference datasets. Using this pre-trained UMAP model, we projected KO and control axial mesoderm data onto the same UMAP 2D embedding space constructed with WT reference data.

Cell density was visualized using a KDE plot. First, we performed random downsampling to adjust the cell number between the LOF control samples. (i) Whole-cell populations were randomly subsampled into a subset to have an equal cell number to the smaller dataset. (ii) Then, axial mesoderm cells were selected and subjected to KDE calculation. KDE was calculated with

the scipy.stat.gaussian_kde function. The calculated KDE was visualized with the matplotlib.pyplot.contour function. We used the same contour threshold levels between the LOF and control samples.

In addition to the UMAP transfer analysis above, the WT data, *lhx1a* crispant and *tyr* crispant data were analysed with UMAP without data transfer (Supplementary Fig. 21). The 10 hpf axial mesoderm cell data were integrated using Seurat functions (SelectIntegrationFeatures(), FindIntegrationAnchors(), and IntegrateData() with default parameters), and then UMAP graph and Louvain cluster were calculated (RunPCA(), FindNeighbors(reduction = "pca", dims = 1:30), RunUMAP(reduction = "pca", dims = 1:30, min.dist = 1), FindClusters(resolution = 1.5)).

NMF

We performed NMF with our *lhx1a* crispants scRNA-seq dataset according to a previous report[32]. The normalized UMI counts were standardized, log-transformed and subjected to NMF calculation with sklearn.decomposition.NMF(n_components=40). Each module was manually annotated by its cluster enrichment and gene ontology calculated with the top 30 genes with high module weight. Gene annotation, weight and ontology are provided in Supplementary Table 3. Gene ontology was calculated with the g:Profiler API (https://biit.cs.ut.ee/gprofiler/page/apis). The background was set to all genes used in the NMF calculation. Clusters governed by a single gene were excluded from our analysis.

Statistical testing

Details of all statistical tests performed are provided in Supplementary Table 4. Scipy stat module (scipy version 1.7.0) was used for statistical analysis. In summary, we selected the statistical method as follows: (i) chi-square test was used to analyse the relationships of categorical variables; (ii) Wilcoxon rank-sum test (Mann–Whitney *U* test) was used when the data distribution type was not apparent; (iii) in cases in which the data distribution followed a Gaussian distribution, a *t*-test was used. Where multiple comparisons were made, the Bonferroni correction was applied.

An alternative hypothesis (one-tailed or two-tailed) was selected depending on the aim of the analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data, including sequencing reads and single-cell expression matrices, are available from the GEO under accession codes GSE72859 (ref. [16]), GSE112824 (ref. [32]) and GSE145298 for the zebrafish profiling from this study; and from ArrayExpress under accession codes E-MTAB-7325 (*Tal1*$^{-/-}$ chimeras) and E-MTAB-7324 (wild-type chimeras). Simulations can be explored at https://celloracle.org.

Code availability

CellOracle code, documentation and tutorials are available at GitHub: https://github.com/morris-lab/CellOracle.

References

1. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).

2. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).

3. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).

4. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).

5. Ji, Y., Lotfollahi, M., Wolf, F. A. & Theis, F. J. Machine learning for perturbational single-cell omics. *Cell Syst.* **12**, 522–537 (2021).

6. Fiers, M. W. E. J. et al. Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genomics* **17**, 246–254 (2018).

7. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

8. Iacono, G., Massoni-Badosa, R. & Heyn, H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* **20**, 110 (2019).

9. Fleck, J. S. et al. Inferring and perturbing cell fate regulomes in human brain organoids. *Nature* https://doi.org/10.1038/s41586-022-05279-8 (2022).

10. Kartha, V. K. et al. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom.* **2**, 100166 (2022).

11. González-Blas, C. B. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Preprint at *bioRxiv* https://doi.org/10.1101/2022.08.19.504505 (2022).

12. Pliner, H. A. et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871 (2018).

13. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).

14. Oki, S. et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, e46255 (2018).

15. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).

16. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).

17. Rekhtman, N., Radparvar, F., Evans, T. & Skoultchi, A. I. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev.* **13**, 1398–1411 (1999).

18. Zhang, P. et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc. Natl Acad. Sci. USA* **96**, 8705–8710 (1999).

19. Nutt, S. L., Metcalf, D., D'Amico, A., Polli, M. & Wu, L. Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. *J. Exp. Med.* **201**, 221–231 (2005).

20. Back, J., Allman, D., Chan, S. & Kastner, P. Visualizing PU.1 activity during hematopoiesis. *Exp. Hematol.* **33**, 395–402 (2005).

21. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl Acad. Sci. USA* **93**, 12355–12358 (1996).

22. Yu, C. et al. Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J. Exp. Med.* **195**, 1387–1395 (2002).

23. Fulkerson, P. C. Transcription factors in eosinophil development and as therapeutic targets. *Front. Med.* **4**, 115 (2017).

24. Hirasawa, R. et al. Essential and instructive roles of GATA factors in eosinophil development. *J. Exp. Med.* **195**, 1379–1386 (2002).

25. Iwasaki, H. et al. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood* **106**, 1590–1600 (2005).

26. Mak, K. S., Funnell, A. P. W., Pearson, R. C. M. & Crossley, M. PU.1 and haematopoietic cell fate: dosage matters. *Int. J. Cell Biol.* **2011**, 808524 (2011).

27. Wontakal, S. N. et al. A large gene network in immature erythroid cells is controlled by the myeloid and B cell transcriptional regulator PU.1. *PLoS Genet.* **7**, e1001392 (2011).

28. Moreau-Gachelin, F., Tavitian, A. & Tambourin, P. *Spi-1* is a putative oncogene in virally induced murine erythroleukaemias. *Nature* **331**, 277–280 (1988).

29. Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* **7**, 105–117 (2007).

30. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).

31. Mikkola, H. K. A. et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia *SCL/tal-1* gene. *Nature* **421**, 547–551 (2003).

32. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

33. Halpern, M. E. Axial mesoderm and patterning of the zebrafish embryo. *Am. Zool.* **37**, 311–322 (1997).

34. Fuentes, R. et al. The maternal coordinate system: molecular-genetics of embryonic axis formation and patterning in the zebrafish. *Curr. Top. Dev. Biol.* **140**, 341–389 (2020).

35. Joyce, A. R. & Palsson, B. Predicting gene essentiality using genome-scale in silico models. *Methods Mol. Biol.* **416**, 433–457 (2008).

36. Hahn, M. W. & Kern, A. D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* **22**, 803–806 (2005).

37. Talbot, W. S. et al. A homeobox gene essential for zebrafish notochord development. *Nature* **378**, 145–149 (1995).

38. Halpern, M. E. et al. Cell-autonomous shift from axial to paraxial mesodermal development in zebrafish floating head mutants. *Development* **121**, 4257–4264 (1995).

39. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).

40. Klatt Shaw, D. et al. Localized EMT reprograms glial progenitors to promote spinal cord repair. *Dev. Cell* **56**, 613–626 (2021).

41. Dal-Pra, S., Thisse, C. & Thisse, B. FoxA transcription factors are essential for the development of dorsal axial structures. *Dev. Biol.* **350**, 484–495 (2011).

42. Shawlot, W. & Behringer, R. R. Requirement for LIml in head-organizer function. *Nature* **374**, 425–430 (1995).

43. Poulain, M. & Lepage, T. Mezzo, a paired-like homeobox protein is an immediate target of nodal signalling and regulates endoderm specification in zebrafish. *Development* **129**, 4901–4914 (2002).

44. Ragvin, A. et al. Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX*, *SOX4*, and *IRX3*. *Proc. Natl Acad. Sci. USA* **107**, 775–780 (2010).

45. Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl Acad. Sci. USA* **101**, 4164–4169 (2004).

46. Lele, Z., Nowak, M. & Hammerschmidt, M. Zebrafish *admp* is required to restrict the size of the organizer and to promote posterior

and ventral development. *Dev. Dyn.* **222**, 681–687 (2001).

47. Lotfollahi, M., Wolf, F. A. & Theis, F. J. scGen predicts single-cell perturbation responses. *Nat. Methods* **16**, 715–721 (2019).

48. Kamimoto, K. et al. Gene regulatory network reconfiguration in direct lineage reprogramming. *Stem Cell Rep.* https://doi.org/10.1016/j.stemcr.2022.11.010 (2022).

49. Shestopalov, I. A., Pitt, C. L. W. & Chen, J. K. Spatiotemporal resolution of the Ntla transcriptome in axial mesoderm development. *Nat. Chem. Biol.* **8**, 270–276 (2012).

50. Bocchi, V. D. et al. The coding and long noncoding single-cell atlas of the developing human fetal striatum. *Science* **372**, eabf5759 (2021).

51. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776 (2010).

52. la Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

53. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).

54. Liu, H. et al. Nonlinear network reconstruction from gene expression data using marginal dependencies measured by DCOL. *PLoS ONE* **11**, e0158247 (2016).

55. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 (2020).

56. Tabula Muris Consortium Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).

57. Liu, C. et al. An ATAC-seq atlas of chromatin accessibility in mouse tissues. *Sci. Data* **6**, 65 (2019).

58. Dahlin, J. S. et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood* **131**, e1–e11 (2018).

59. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

60. Weinreb, C., Wolock, S. & Klein, A. M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018).

61. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **9**, e98679 (2014).

62. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).

63. Klein, C., Marino, A., Sagot, M.-F., Vieira Milreu, P. & Brilli, M. Structural and dynamical analysis of biological networks. *Brief. Funct. Genomics* **11**, 420–433 (2012).

64. Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein *tal-1*/*SCL*. *Nature* **373**, 432–434 (1995).

65. Robb, L. et al. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc. Natl Acad. Sci. USA* **92**, 7075–7079 (1995).

66. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

67. Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).

68. Buenrostro, J. D. et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell* **173**, 1535–1548 (2018).

69. Burda, P., Laslo, P. & Stopka, T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia* **24**, 1249–1257 (2010).

70. Galloway, J. L., Wingert, R. A., Thisse, C., Thisse, B. & Zon, L. I. Loss of Gata1 but not Gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos. *Dev. Cell* **8**, 109–116 (2005).

71. Stopka, T., Amanatullah, D. F., Papetti, M. & Skoultchi, A. I. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *EMBO J.* **24**, 3712–3723 (2005).

72. Hyung, C. S. et al. C/EBPα determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation. *Blood* **107**, 4308 (2006).

73. Vassen, L., Dührsen, U., Kosan, C., Zeng, H. & Möröy, T. Growth factor independence 1 (Gfi1) regulates cell-fate decision of a bipotential granulocytic-monocytic precursor defined by expression of Gfi1 and CD48. *Am. J. Blood Res.* **2**, 228 (2012).

74. van der Meer, L. T., Jansen, J. H. & van der Reijden, B. A. Gfi1 and Gfi1b: key regulators of hematopoiesis. *Leukemia* **24**, 1834–1843 (2010).

75. Inoue, A. et al. Elucidation of the role of LMO2 in human erythroid cells. *Exp. Hematol.* **41**, 1062–1076 (2013).

76. Ben-David, Y., Gajendran, B., Sample, K. M. & Zacksenhaus, E. Current insights into the role of Fli-1 in hematopoiesis and malignant

transformation. *Cell. Mol. Life Sci.* **79**, 163 (2022).

77. Masuya, M. et al. Dysregulation of granulocyte, erythrocyte, and NK cell lineages in *Fli-1* gene-targeted mice. *Blood* **105**, 95–102 (2005).

78. Hoshijima, K. et al. Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Dev. Cell* **51**, 645–657 (2019).

79. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).

80. Peterson, S. M. & Freeman, J. L. RNA Isolation from embryonic zebrafish and cDNA synthesis for gene expression analysis. *J. Vis. Exp.* **30**, e1470 (2009).

81. Fürthauer, M., Thisse, B. & Thisse, C. Three different *noggin* genes antagonize the activity of bone morphogenetic proteins in the zebrafish embryo. *Dev. Biol.* **214**, 181–196 (1999).

82. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2007).

83. Alles, J. et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 44 (2017).

Acknowledgements

We thank members of the S.A.M. laboratory for discussions; J. Magee and T. Druley for critical feedback; A. Krezel for help with gRNA design; M. Ryan and T. Tsai for advice on HCR; and D. Klatt Shaw for advice on CRISPR–Cas9 gene disruption. This work was funded by the National Institute of General Medical Sciences grant R01 GM126112 and Silicon Valley Community Foundation, Chan Zuckerberg Initiative grants HCA2-A-1708-02799 and DAF2021-238797 (both to S.A.M.), and by grant R35 GM118179 to L.S.-K. S.A.M. is supported by an Allen Distinguished

Investigator Award (through the Paul G. Allen Frontiers Group), a Vallee Scholar Award, a Sloan Research Fellowship and a New York Stem Cell Foundation Robertson Investigator Award; K.K. is supported by a Japan Society for the Promotion of Science Postdoctoral Fellowship; B.S. is supported by a postdoctoral fellowship from the Washington University in St Louis Center of Regenerative Medicine; and C.M.H. is supported by a National Science Foundation Graduate Research Fellowship (DGE-2139839 and DGE-1745038).

Author information

Authors and Affiliations

1. Department of Developmental Biology, Washington University School of Medicine in St Louis, St Louis, MO, USA

   Kenji Kamimoto, Blerta Stringa, Christy M. Hoffmann, Kunal Jindal, Lilianna Solnica-Krezel & Samantha A. Morris

2. Department of Genetics, Washington University School of Medicine in St Louis, St Louis, MO, USA

   Kenji Kamimoto, Christy M. Hoffmann, Kunal Jindal & Samantha A. Morris

3. Center of Regenerative Medicine, Washington University School of Medicine in St Louis, St Louis, MO, USA

   Kenji Kamimoto, Blerta Stringa, Christy M. Hoffmann, Kunal Jindal, Lilianna Solnica-Krezel & Samantha A. Morris

Contributions

K.K. and S.A.M. conceived the research. K.K. led the computational and experimental work, assisted by C.M.H. and K.J. and supervised by S.A.M. B.S. performed the zebrafish experimental validations, supervised by L.S.-K. All authors participated in the interpretation of data and writing.

Corresponding author

Correspondence to Samantha A. Morris.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Carl de Boer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Overview of the CellOracle workflow.

(a) Overview of the CellOracle context-dependent GRN model construction method. First, genomic DNA sequence and TF-binding-motif information provide all potential regulatory links to construct a 'base GRN.' CellOracle uses scATAC-seq data to identify accessible promoter and enhancer DNA sequences in this step. The DNA sequence of these regulatory elements is scanned for TF-binding motifs, generating a list of potential regulatory connections between a TF and its target genes (left). Next, active connections (described below), dependent on cell state or cell type, are identified from all potential connections in the base GRN. CellOracle builds

machine-learning (ML) models for this step that predict the quantitative relationship between the TF and the target gene. The ML model fitting results present the certainty of connection as a distribution, enabling the identification of GRN configurations by removing inactive connections from the base GRN structure. **(b—d)** Overview of signal propagation simulation. CellOracle leverages an inferred GRN model to simulate how target gene expression changes in response to the changes in regulatory gene expression. **(b)** The input TF perturbation (shown in yellow) is propagated side-by-side within the network model. **(c)** Input data and GRN coefficient matrix format used in the signal propagation calculation. **(d)** Leveraging the linear predictive ML algorithm features, CellOracle uses the GRN model as a function to perform the signal propagation calculation. Iterative matrix multiplication steps enable the estimation of indirect and global downstream effects resulting from the perturbation of a single TF. **(e)** After signal propagation, the simulated gene expression shift vector is converted into a 2D vector and projected onto the low-dimensional space. Details are described in the Methods.

Extended Data Fig. 2 Benchmarking of inferred GRN configurations.

(a,b) We benchmarked the CellOracle GRN modelling method against pre-existing GRN inference algorithms: WGCNA, DCOL, GENIE3, and SCENIC. Details of input data and ground-truth data are described in the Methods. We generated a base GRN using the Cusanovich mouse sci-ATAC-seq atlas dataset[13] or UCSC mm9 promoter DNA sequence data. CellOracle scored better than or comparable to other algorithms. CellOracle results with a promoter base GRN received lower but comparable scores than the scATAC-seq base GRN results. In addition, we tested the CellOracle GRN method using two impaired base GRN datasets (Scrambled motif base GRN and no base GRN) to investigate how the base GRN data contributes to its performance. **(a)** AUROC (Area Under the Receiver Operating Characteristic curve) heat map. The top score in each condition is highlighted with a red rectangle. **(b)** EPR (Early Precision Ratio) heat map. EPR represents the EP ratio relative to the random model ER score. An EPR of less than 1 indicates that the GRN inference results are no better than random prediction. **(c,d)** The performance of CellOracle

was tested after downsampling cells. GRN models were made after downsampling to 400, 200, 100, 50, 25, and 10 cells. We recommend at least 50 cells for GRN inference based on these results. CellOracle used the same mouse scATAC-seq base GRN as **a** and **b**. The Liver_2 sample contains less than 400 cells. **(e,f)** GRN inference performance comparison between different base GRN data generated from various tissue types. The top score in each condition is highlighted with a red rectangle.

Extended Data Fig. 3 CellOracle analysis of Paul et al. haematopoiesis data.

(a) Force-directed graph of 2,730 myeloid progenitor cells from Paul et al.[16] with all clusters labelled. DC = Dendritic Cell; Ery = Erythrocyte; GMP = Granulocyte–Monocyte Progenitor; Gran = Granulocyte; Lym = Lymphoid; MEP = Megakaryocyte–Erythrocyte Progenitor; Mk = Megakaryocyte; Mo = Monocyte. We removed the DC and Lymphoid cell clusters to focus on myeloid cell differentiation. **(b)** Degree distribution of the MEP_0 cluster GRN model. After making the GRN model for each cluster, network edges were pruned. Then, we counted the network degree (k), representing the number of network edges for each gene. P(k) is the frequency of network degree k. The relationship between k and P(k) was visualized after log transformation to test whether the data follow a power law, in which there is a linear relationship between log(k) and log(P(k)). The R-squared value (R^2) was calculated to quantify the degree of the linear relationship. The same analysis was performed on the randomized GRN (lower panel). **(c)** Top 30 genes ranked by degree centrality in the MEP_0 cluster GRN. **(d)** *Gata1* gene expression (log-transformed UMI) projected onto the force-directed graph (left) and violin plot grouped by cell-type annotation (right). **(e)** *Spi1* gene expression (log-transformed UMI) projected onto the force-directed graph (left) and violin plot grouped by cell-type annotation (right). **(f)** Systematic KO simulation of TFs in the GM (Granulocyte–Monocyte) and ME (Megakaryocyte–Erythrocyte) lineages. The sum of the negative perturbation scores is calculated for each TF to quantify the perturbation effect along each lineage. **(g)** Negative PS sum cut-off value calculation. Cut-off values were calculated for GM and ME lineage simulations based on the distribution of PS sum score calculated from the randomized simulation result (false-positive rate (FPR) = 0.01).

Extended Data Fig. 4 Perturbation score calculation and interpretation.

(a—d) Schematic for perturbation score (PS) calculation. CellOracle calculates a PS by comparing the direction of the simulated cell state transition with the direction of cell differentiation. **(a)** Schematic for differentiation vector calculations. First, the pseudotime data are summarized by grid points. Then, CellOracle calculates a 2D gradient vector of the pseudotime data representing the directionality of differentiation pseudotime. **(b)** Calculation of the inner-product value between the differentiation vector and gene perturbation vectors. First, the results of the perturbation simulation are converted into the same vector field format as the differentiation vector field, and the inner product of these vectors is calculated to produce a PS. **(c)** A positive PS (magenta) suggests the perturbation vector and differentiation vector share a similar direction, thus, suggesting the TF perturbation would promote differentiation. In contrast, a negative PS (green) represents inhibited differentiation. **(d)** Schematic for perturbation score interpretation. A positive perturbation score (green) predicts that the perturbation promotes differentiation. A negative perturbation score (purple) represents inhibited differentiation.

Extended Data Fig. 5 CellOracle TF KO simulation results for Paul et al. haematopoiesis data: part 1.

a–**d**, CellOracle KO simulation for four key TF regulators of haematopoiesis: *Klf1* (**a**), *Gfi1b* (**b**), *Gfi1* (**c**) and *Irf8* (**d**) reported in[15,29]. The simulated cell state transition vector field is visualized with perturbation scores (PS; magenta: negative score; green: positive score). The right column shows a summary of the TF role based on the CellOracle simulation results, cell transition vector, and PS. For example, a positive PS in the TF KO simulation (green) implies that the TF has a role in cell state maintenance or inhibiting cell differentiation. In contrast, a negative PS in the KO simulation (magenta) implies that the TF normally promotes cell differentiation.

Extended Data Fig. 6 CellOracle TF KO simulation results for Paul et al. haematopoiesis data: part 2.

a–**d**, CellOracle KO simulation results for *Gata2* (**a**), *Runx1* (**b**), *Fli1* (**c**) and *Lmo2* (**d**). The simulated cell state transition vector field is visualized with perturbation scores (PS; magenta: negative score; green: positive score). The right column shows a summary of the TF role based on the CellOracle simulation results, cell transition vector, and PS. For example, a positive PS in the TF KO simulation (green) implies that the TF has a role in cell state maintenance or inhibiting cell differentiation. In contrast, a negative PS in the KO simulation (magenta) implies that the TF normally promotes cell differentiation.

Extended Data Fig. 7 Dahlin et al. mouse haematopoiesis scRNA-seq data.

a, Force-directed graph of 44,082 myeloid progenitor cells from Dahlin et al.[58] with all clusters labelled. MPP = Multipotent Progenitor; GMP = Granulocyte–Monocyte Progenitor; Gran = Granulocyte; LP = Lymphoid progenitor; MEP = Megakaryocyte–Erythrocyte Progenitor; Mk = Megakaryocyte; Mo = Monocyte; Baso = Basophil. **(b)** Marker gene expression (log-transformed UMI) projected onto the force-directed graph. *Procr* = MPP marker; *Epor* = Erythrocyte marker; *Itga2b* = Mk marker; *Flt3* = LP marker; *Mpo* = Gran/Mo marker; *Ms4a2* = Baso marker. **(c)** Pseudotime values projected onto the force-directed graph. **(d)** Differentiation vector calculated from the pseudotime gradient. ME and GM lineages are highlighted. **(e)** *Csf1r* and *Cebpε* gene expression projected onto the force-directed graph. The right panel is a magnified area of the GM lineage. *Csf1r* is a monocyte marker, and *Cebpε* is a granulocyte marker. **(f)** Early lineage bifurcation between monocytes and granulocytes is observed on the force-directed graph.

Extended Data Fig. 8 Setty et al. human haematopoiesis scRNA-seq data.

(a) Force-directed graph of 5,610 myeloid progenitor cells from Setty et al.[67] with all clusters labelled. MPP = Multipotent Progenitor; GMP = Granulocyte–Monocyte Progenitor; Gran = Granulocyte; MEP = Megakaryocyte–Erythrocyte Progenitor; Mk = Megakaryocyte. **(b)** Marker gene expression (log-transformed UMI) projected onto the force-directed graph. *PROCR* = MPP marker; *EPOR* = Erythrocyte marker; *ITGA2B* = Mk marker; *FLT3* = LP marker; *MPO* = Gran/Mo marker; *MS4A2* = Baso marker. **(c)** Pseudotime values projected onto the force-directed graph. **(d)** Differentiation vector calculated from the pseudotime gradient. ME and GM lineages are highlighted. **(e)** *CSF1R* and *CEBPE* gene expression projected onto the force-directed graph. The right panel is a magnified area of the GM lineage. The *CSF1R* is a monocyte marker, and *CEBPE* is a granulocyte marker. **(f)** Early lineage bifurcation between monocytes and granulocytes is observed on the force-directed graph.

Extended Data Fig. 9 CellOracle validation using experimentally measured cell density in *Tal1* KO in Pijuan-Sala et al. gastrulation and organogenesis scRNA-seq data.

(a) UMAP plot of chimeric E8.5 embryos of wild-type (WT) and *Tal1* KO cells (25,307 cells and 26,311 cells, respectively) from a published scRNA-seq atlas of mouse gastrulation and organogenesis[30]. **(b)** *Tal1* gene expression (log-transformed UMI) projected onto the UMAP plot. **(c)** Pseudotime gradient vector field used in the perturbation score (PS) calculations. Developmental pseudotime was calculated using the DPT method with WT chimera scRNA-seq data and then converted into a 2D gradient vector field. **(d)** PS and cell transition vector field of the *Tal1* KO simulation. **(e)** The magnified area of erythrocyte differentiation predicts inhibition or arrest of cell differentiation at the haematoendothelial progenitor stage. **(f)** The Markov random walk simulation result predicts high cell density in the haematoendothelial progenitor cluster and lower cell density at later stages, indicating that *Tal1* KO would induce differentiation arrest at the haematoendothelial progenitor stage. **(g)** Experimentally measured *Tal1* KO data. The kernel cell density of whole chimera (left), WT (middle), and *Tal1* KO cells (right) were calculated after downsampling each condition (25,307 cells) to control for sample size. A scatter plot of

whole chimera cells is shown as background (light grey) to highlight the overall cell trajectory structure. **(h)** The bar plot shows the cell type composition in each sample (right panel). Overall, the experimental result aligns with the simulated predictions. The relative fold change between WT and KO samples is also shown in Supplementary Table 4. **(i)** Perturbation score and cell transition vector field of the *Tal1* conditional KO simulation in the erythroid lineage. *Tal1* expression was set to zero in the Blood progenitor and Erythrocyte clusters; CellOracle simulates KO effects in later erythroid differentiation stages. **(j)** The Markov simulation result shows uniform cell density, predicting that *Tal1* KO would not induce differentiation arrest in a conditional KO targeting later stages of erythroid differentiation.

Extended Data Fig. 10 CellOracle *noto* LOF simulation with Farrell et al. zebrafish embryogenesis data.

(a) 2D force-directed graph of a published atlas[32] of zebrafish embryogenesis (n = 25,711 cells). **(b)** Main trajectory partitioned into four sub-branches. **(c)** Bar plots depicting the number of TFs after variable gene selection (black), the number of TFs with >1 network edge in the inferred GRN model (dark grey), and the number of TFs expressed in >1% of cells (light grey). **(d)** CellOracle *noto* LOF simulation result (left) and simulation results with a randomized GRN model (right) for the notochord lineage. Simulated cell state transitions for each cell were converted to a vector field and visualized with a scatter plot (shown in grey). **(e)** *Noto* LOF simulation for the prechordal plate lineage. **(f)** CellOracle *noto* LOF simulation vector is shown at single-cell resolution. Cells in the Notochord cluster are shown in orange, while the Prechordal Plate cells are shown in blue. The right panel is the magnified area. **(g)** Force-directed graph of the Other mesendoderm sub-branch with cell cluster annotations from the Farrell et al. study[32] (n = 10,265 cells). **(h)** Pseudotime data are projected onto the force-directed graph. **(i)** The Somite lineage, defined in the previous Farrell et al. study[32], is in red. **(j)** Pseudotime gradient vector field calculated for the Somite lineage. **(k)** *Noto* LOF simulation vector field in the cells of the Somite lineage are shown with perturbation scores.

[Extended Data Fig. 11 Zebrafish scRNA-seq experiments for *noto* LOF analysis.]

(a) Schematic illustration of zebrafish scRNA-seq experiments. (1) The reference dataset was generated using cells from 6, 8, and 10 hpf wild-type (WT) embryos. To assess *noto* LOF, we also assayed (2) $flh^{n1/n1}$ mutants and (3) *noto/flh* crispants at 10 hpf (~25 embryos per sample; Methods). **(b)** Cell cluster composition comparing *tyr* crispant (control) with WT cells, showing similar cell distributions. After data integration, cell-type labels were transferred from the whole WT 6, 8, and 10 hpf reference data (see Methods). **(c)** Sample label projected onto the *t*-SNE plot. $flh^{n1/n1}$ mutant and control sample (left, n = 57,175 cells, 2 independent biological replicates for each sample), and *t*-SNE plot of *noto* crispant and *tyr* crispant samples (right, n = 9,185 cells, 2 biological, 3 technical replicates for *noto* crispant; n = 46,440 cells, n = 3 independent biological, 5 technical replicates for *tyr* crispant). **(d)** Cluster annotation label projected onto the *t*-SNE plot. WT zebrafish cells (left, n = 38,606 cells, two technical replicates per stage), $flh^{n1/n1}$ mutant and control sample (middle), *noto* crispant and *tyr* crispant samples (right). **(e)** Cell cluster composition comparing LOF samples with the control samples.

[Extended Data Fig. 12 Zebrafish notochord regulator screening with CellOracle and initial experimental validation.]

(a) Overview of the systematic LOF simulation and quantification method. CellOracle LOF simulation was performed for 232 TFs in the Notochord lineage to calculate the perturbation score (PS). The sum of the negative PS was calculated for each TF in the selected area between digitized pseudotime 0 to 3, before lineage specification. **(b)** Degree centrality score in the Notochord cluster GRN (left), gene expression specificity score in the Axial mesoderm sub-branch (middle), and mean expression value in the Axial mesoderm sub-branch (right) were calculated for the top 30 TFs selected in the systematic simulation to further prioritize candidate genes for experimental validation. We selected genes in the top 50% of these scores. Please refer to the Methods for the detailed selection procedure. We selected three candidates for experimental validation: *lhx1a*, *sebox*, and

irx3a. (**c**,**e**,**g**) Cell cluster composition in axial mesoderm cells, comparing LOF (*lhx1a*, *sebox*, and *irx3a)* samples with control samples. Cell cluster composition comparison was performed with a Chi-square test, Two-tailed Bonferroni correction. *lhx1a* experiment: Early axial mesoderm p = 0.000229717, Early Notochord $p = 1.08\times10^{-21}$, Notochord $p = 4.38\times10^{-6}$, Prechordal Plate $p = 1.42\times10^{-10}$. *Sebox* experiment: Early axial mesoderm $p = 3.01\times10^{-6}$, Early Notochord $p = 2.87\times10^{-6}$, Notochord $p = 4.38\times10^{-6}$, Prechordal Plate $p = 4.17\times10^{-9}$. The left panels show cluster composition in the merged data, and the right panels show individual scRNA-seq batch. *lhx1a* LOF produced the most significant changes in cell composition. (**d**,**f**,**h**) Comparison of notochord marker gene expression between LOF and control samples. scRNA-seq gene expression in the Notochord lineage clusters is shown as a violin plot. Late-stage notochord markers, *twist2* and *nog1*, or broad/early notochord markers, *noto* and *tbxta*, are visualized. Statistical tests: Wilcoxon rank-sum test, two-tailed with Bonferroni p-value correction. *lhx1a* experiment: *twist2* $p = 7.118\times10^{-64}$, *nog1* $p = 7.757\times10^{-67}$, *noto* $p = 7.718\times10^{-11}$. *sebox* experiment: *twist2* $p = 8.022\times10^{-10}$, *nog1* $p = 3.184\times10^{-3}$, *tbxta* $p = 1.551\times10^{-3}$. *irx3a* experiment: *twist2* p = 0.000012. **(c)** n = 720 cells and 1,686 cells for *lhx1a* crispant and *tyr* crispant, respectively. **(e)** n = 1,216 cells and 1,703 cells for *sebox* crispant and *tyr* crispant, respectively. **(g)** n = 1,176 cells and 1,651 cells for *irx3a* crispant and *tyr* crispant, respectively.

Extended Data Fig. 13 Zebrafish scRNA-seq experiments for *lhx1a* LOF analysis.

(**a**,**b**) *t*-SNE plot of *lhx1a* crispant (n = 45,582 cells, 4 biological replicates) and *tyr* control crispant samples (n = 76,163 cells, 5 biological, 7 technical replicates). **(a)** Cluster annotation labels transferred from WT reference data projected onto the *t*-SNE plot. **(b)** Sample label projected onto the *t*-SNE plot. **(c)** Cell cluster composition comparing *lhx1a* crispant and *tyr* control crispant samples as a proportion of cells from the whole embryo. **(d)** Cell density in the axial mesoderm is visualized as a kernel cell density contour plot. The cell number is downsampled to match the cell number before kernel cell density calculation (n = 260, 290, 336, and 367 for *lhx1a* crispant 1~4, n = 248, 234, 344, 316, 213, 286, and 350 for *tyr* crispant 1~7). The

same contour threshold values are used for the visualization. **(e)** Cell cluster composition in the axial mesoderm clusters comparing *lhx1a* crispant and *tyr* control crispant samples. The left panels show cluster composition in the merged data, while the right panels show the individual scRNA-seq batch. **(f)** The top 30 NMF module weights for the Early notochord module (left) and the Late notochord module (right) are shown as a bar plot.

Supplementary information

Supplementary Information

This file contains Supplementary Figures 1–26, Supplementary Tables 1–6 and Supplementary References

Reporting Summary

Supplementary Table 1

Details of CellOracle pre-built base GRN data. This table details the different species for which we have constructed base GRNs.

Supplementary Table 2

TF annotation and previous study references. This table details the previously reported knockout phenotypes in mouse haematopoiesis and zebrafish development.

Supplementary Table 3

Negative PS sum scores for zebrafish KO simulation. This table details the ranked list of predicted regulators of somitogenesis, with published phenotypes.

Supplementary Table 4

Details of statistical tests. This table details the statistical tests used throughout this study.

Supplementary Table 5

Lhx1a crispant NMF data. This table details the NMF analysis of *Lhx1a* crispants. For statistical tests, Wilcoxon rank-sum test, Two-tailed was used. The p-value was corrected with the Bonferroni correction method.

Supplementary Table 6

sgRNA and primer sequences. This table details the sgRNA and primer sequences used in the zebrafish experiments.

Supplementary Video 1

Visualization of the differentiation vector field and TF KO vector fields. The same 2D vector field used in Fig. 1c–f is visualized as a movie. The left panel shows the differentiation vector field of Paul et al. haematopoiesis data. The centre and right panels show CellOracle KO simulation vectors for *Spi1* and *Gata1* transcription factors, respectively. The particle visual effects were generated with Unity (version 2020.3.0f1) and Visual Effect Graph Asset: https://docs.unity3d.com/Packages/com.unity.visualeffectgraph@13.1/manual/VisualEffectGraphAsset.html

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the

article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Kamimoto, K., Stringa, B., Hoffmann, C.M. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023). https://doi.org/10.1038/s41586-022-05688-9

- Received: 04 January 2022
- Accepted: 28 December 2022
- Published: 08 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05688-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

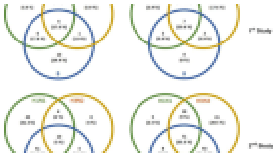
Provided by the Springer Nature SharedIt content-sharing initiative

An oracle predicts regulators of cell identity

- Jeffrey A. Farrell

Nature News & Views 08 Feb 2023

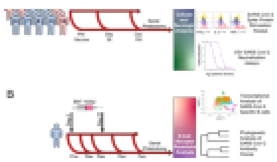
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05688-9



Distinct immunological and molecular signatures underpinning influenza vaccine responsiveness in the elderly

12 November 2022

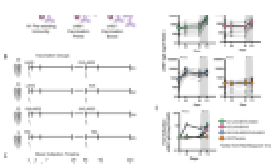
Peggy Riese, Stephanie Trittel, … Carlos A. Guzmán



Single-cell profiling of the antigen-specific response to BNT162b2 SARS-CoV-2 RNA vaccine

16 June 2022

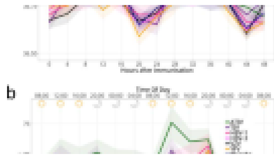
Kevin J. Kramer, Erin M. Wilfong, … Jonathan M. Irish



Transcriptome signatures preceding the induction of anti-stalk antibodies elicited after universal influenza vaccination

10 December 2022

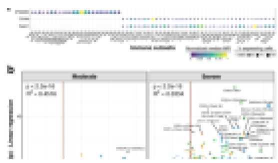
Teresa Aydillo, Ana S. Gonzalez-Reiche, … Adolfo García-Sastre



[Characterization of potential biomarkers of reactogenicity of licensed antiviral vaccines: randomized controlled clinical trials conducted by the BIOVACSAFE consortium]

30 December 2019

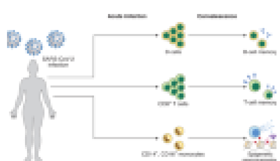
January Weiner, David J. M. Lewis, … Giuseppe Del Giudice



[Immune phenotypes that are associated with subsequent COVID-19 severity inferred from post-recovery samples]

25 November 2022

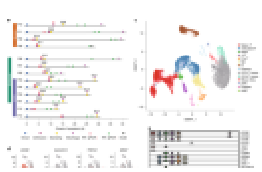
Thomas Liechti, Yaser Iftikhar, … Mario Roederer



[Immune memory in individuals with COVID-19]

09 June 2021

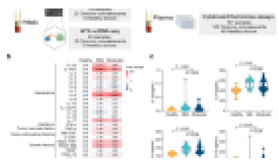
Mihai G. Netea & Yang Li



[Single-cell landscape of immunological responses in patients with COVID-19]

12 August 2020

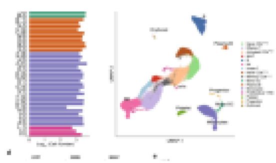
Ji-Yuan Zhang, Xiang-Ming Wang, … Fu-Sheng Wang



[Innate immune imprints in SARS-CoV-2 Omicron variant infection convalescents]

16 November 2022

Zhiqing Li, Xiaosu Chen, … Xuetao Cao



[Single-cell immune profiling reveals distinct immune response in asymptomatic COVID-19 patients]

16 September 2021

Xiang-Na Zhao, Yue You, … Mai-Juan Ma

- Article
- [Published: 04 January 2023]

Influenza vaccination reveals sex dimorphic imprints of prior mild

COVID-19

- Rachel Sparks,
- William W. Lau,
- Can Liu,
- Kyu Lee Han,
- Kiera L. Vrindten,
- Guangping Sun,
- Milann Cox,
- Sarah F. Andrews,
- Neha Bansal,
- Laura E. Failla,
- Jody Manischewitz,
- Gabrielle Grubbs,
- Lisa R. King,
- Galina Koroleva,
- Stephanie Leimenstoll,
- LaQuita Snow,
- OP11 Clinical Staff,
- Jinguo Chen,
- Juanjie Tang,
- Amrita Mukherjee,
- Brian A. Sellers,
- Richard Apps,
- Adrian B. McDermott,
- Andrew J. Martins,
- Evan M. Bloch,
- Hana Golding,
- Surender Khurana &
- …
- John S. Tsang

Show authors

Nature **volume 614**, pages 752–761 (2023)

- 24k Accesses
- 1 Citations
- 393 Altmetric
- Metrics details

Abstract

Acute viral infections can have durable functional impacts on the immune system long after recovery, but how they affect homeostatic immune states and responses to future perturbations remain poorly understood[1,2,3,4]. Here we use systems immunology approaches, including longitudinal multimodal single-cell analysis (surface proteins, transcriptome and V(D)J sequences) to comparatively assess baseline immune statuses and responses to influenza vaccination in 33 healthy individuals after recovery from mild, non-hospitalized COVID-19 (mean, 151 days after diagnosis) and 40 age- and sex-matched control individuals who had never had COVID-19. At the baseline and independent of time after COVID-19, recoverees had elevated T cell activation signatures and lower expression of innate immune genes including Toll-like receptors in monocytes. Male individuals who had recovered from COVID-19 had coordinately higher innate, influenza-specific plasmablast, and antibody responses after vaccination compared with healthy male individuals and female individuals who had recovered from COVID-19, in part because male recoverees had monocytes with higher IL-15 responses early after vaccination coupled with elevated prevaccination frequencies of 'virtual memory'-like $CD8^+$ T cells poised to produce more IFNγ after IL-15 stimulation. Moreover, the expression of the repressed innate immune genes in monocytes increased by day 1 to day 28 after vaccination in recoverees, therefore moving towards the prevaccination baseline of the healthy control individuals. By contrast, these genes decreased on day 1 and returned to the baseline by day 28 in the control individuals. Our study reveals sex-dimorphic effects of previous mild COVID-19 and suggests that viral infections in humans can establish new immunological set-points that affect future immune responses in an antigen-agnostic manner.

Main

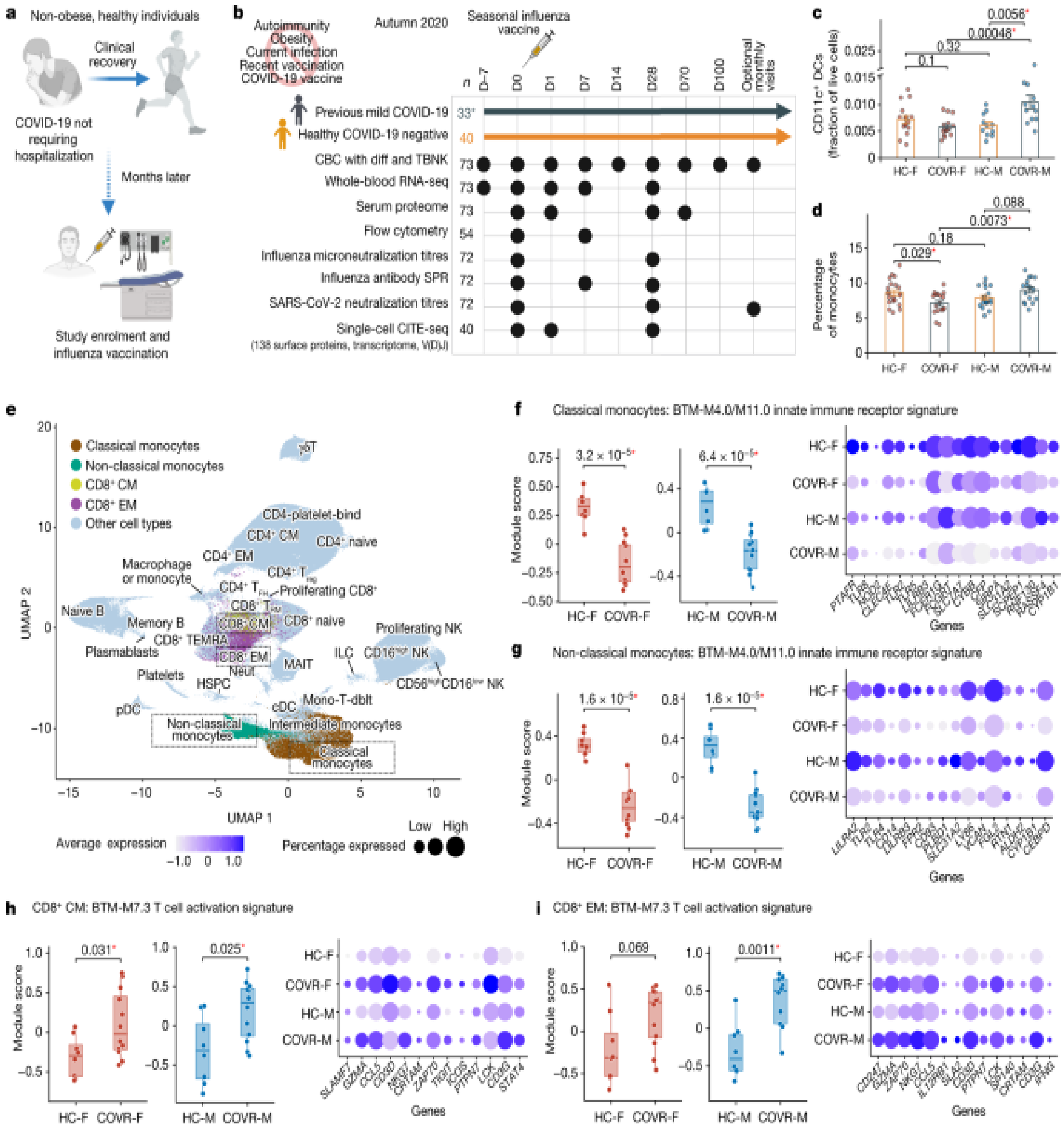
Examples of long-term immunological effects of both chronic and resolved viral infections have been described; for example, after recovery from natural acute measles infection, there is marked reduction in humoral immunity and increased susceptibility to non-measles infections for months to years[1]. Live vaccines such as Bacillus Calmette–Guérin (BCG) and measles can impart ‘training’ effects on innate immune cells such as monocytes and their long-lived progenitors, which could underlie the pathogen non-specific effects of BCG in reducing all-cause mortality in infants[5,6]. COVID-19 can result in persistent clinical sequelae for months after infection, both in hospitalized and mild cases[7]. Although the spectrum of clinical manifestations of post-acute COVID-19 syndrome (also known as long COVID) is expanding, our understanding of the molecular and cellular immunological changes after recovery from SARS-CoV-2 infection is lacking. A better understanding of the functional immune imprints of mild COVID-19 might have particularly important public health implications given that this population constitutes most COVID-19 recoverees. More broadly, the fundamental issues of whether and how homeostatic baseline immune states may have been altered by viral infections, and whether any such alterations may affect responses to future challenges (such as infection or vaccination, with shared or distinct antigens) remain poorly understood.

Here we took advantage of a unique opportunity and epidemiological environment during the early fall of 2020, months after the first wave of COVID-19, when those with mild COVID-19 had recovered clinically, but before they could be reinfected by SARS-CoV-2 or receive COVID-19 vaccination (which was not available until late 2020); moreover, the prevalence of other respiratory infections was extremely low during this time[8]. We enrolled and comparatively assessed healthy individuals who (1) recovered from non-hospitalized, mild cases of COVID-19 and (2) age- and sex-matched controls who never had COVID-19, all from the same geographical region. In addition to assessing the post-COVID-19 immunological statuses, we used influenza vaccination to evaluate the immune responses of these two populations at the serological, transcriptional, proteomic and cellular levels. These analyses reveal basic principles regarding what happens to the immune system after two well-

defined immunological encounters in humans: mild COVID-19 as a natural infectious perturbation and influenza vaccination as a controlled and timed intervention with non-SARS-CoV-2 antigens.

Individuals with previous symptomatic SARS-CoV-2 infection ($n = 31$; diagnosed by nasal PCR test) or asymptomatic infection ($n = 2$; diagnosed by antibody test; Methods), and age- and sex-matched healthy control individuals ($n = 40$) with no history of COVID-19 (and negative by antibody test) were recruited from the community during the fall of 2020 and followed longitudinally (Fig. 1a and Methods). The average time after COVID-19 diagnosis was 151 days for recoverees (Extended Data Fig. 1a and Extended Data Table 1) who had clinically mild illness during acute disease that did not require hospitalization (self-reported average length of illness, 16.5 days) and no major medical comorbidities, including infection at the time of enrolment, obesity (body mass index > 30) or autoimmune disease (Fig. 1b). None of the participants was enrolled in COVID-19 vaccine trials, nor did they receive recent vaccination of any kind before administration of the seasonal influenza vaccine in this study. A small number of individuals continued to have mild self-reported sequelae from their illness at study enrolment (3 males and 8 females), the most common being loss of taste and/or smell (Extended Data Table 1). Female participants were more likely to have sequelae (Fisher's exact test, $P = 0.09$ for all participants, $P = 0.03$ for those aged <65 years), at a rate similar to that reported in other large studies[9].

Fig. 1: Study overview and baseline differences.



a, Schematic of the study concept and design. **b**, Data generated in the study. Both participants who had recovered from COVID-19 and healthy control individuals were enrolled at seven days before vaccination (D−7) and were sampled at the indicated timepoints relative to the day of influenza vaccination. The number of participants assayed for each data type is indicated. CBC with diff, complete blood count with differential; SPR, surface plasmon resonance; TBNK, T and B lymphocyte and natural killer cell phenotyping. Where indicated by an asterisk (*), two asymptomatic individuals (based on antibodies) were included. **c**, Comparison of the

proportion of $CD11c^+$ DCs (as the fraction of live cells from flow cytometry) between the COVR-F ($n = 15$), HC-F ($n = 16$), COVR-M ($n = 12$) and HC-M ($n = 11$) groups at D0. The error bars indicate the s.e.m. of each group. **d**, Similar to **c**, but for monocytes (from CBC; *y* axis) between the COVR-F ($n = 17$), COVR-M ($n = 16$), HC-F ($n = 21$) and HC-M ($n = 19$) groups at the baseline (average of D−7 and D0). **e**, Uniform manifold approximation and projection (UMAP) analysis of the CITE-seq single-cell data showing clustering of cells on the basis of the expression of cell-surface protein markers (632,100 single cells from all timepoints with CITE-seq data: days 0, 1, 28). The coloured and boxed cell clusters are further examined in **f**–**i**. CD4-platelet-bind: $CD4^+$ T cells with platelet markers; CM, central memory; ILC, innate lymphoid cells; Mono-T-dblt, monocytes and T cell doublets; T_{FH}, T follicular helper cells; T_{reg}, regulatory T cells; T_{RM}, tissue resident memory T cells; HSPC, haematopoietic stem and progenitor cell; Neut, neutrophils; pDC, plasmacytoid dendritic cells; cDC, conventional dendritic cells; MAIT: mucosal-associated invariant T cells. **f**, Comparison of the innate immune receptor (IIR) signature scores (Methods) between the HC-F ($n = 8$) and COVR-F ($n = 12$) (left box) and HC-M ($n = 8$) and COVR-M ($n = 12$) (right box) groups using the CITE-seq classical monocyte pseudobulk expression data at D0 (left). Each point represents a participant. Right, the average gene expression of selected genes, including those in the Gene Ontology (GO) pattern recognition receptor activity and immune receptor activity gene sets. **g**, Similar to **f**, but showing the non-classical monocyte population at D0. **h**, Similar to **f**, but showing the T cell activation (BTM-M7.3) module scores of $CD8^+$ CM T cells at D0. The average gene expression of the selected leading-edge genes shared by male and female from the gene set enrichment analysis (GSEA) is shown (Methods). **i**, Similar to **h**, but showing the $CD8^+$ EM T cell population at D0. All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × interquartile range (IQR) from box limits in each direction (upper and lower whiskers). Unless otherwise noted, statistical significance of difference between groups was determined using two-tailed Wilcoxon rank-sum tests. Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagrams in **a** and **b** were created using BioRender.

Baseline of mild COVID-19 recoverees

Longitudinal multi-omics profiling was performed using whole-blood transcriptomics (WBT) analysis, single-cell analysis of 138 surface proteins, transcriptome and V(D)J sequence analysis using cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)[10], serum protein profiling, antibody characterization, peripheral blood immune cell frequencies with haematological parameters from a complete blood count (CBC), as well as clinical and research flow cytometry covering major immune cell lineages and subsets (Fig. 1b and Supplementary Fig. 1). We first assessed the baseline prevaccination differences between the recoverees and the age- and sex-matched healthy control individuals. As sex-dependent immune responses to COVID-19 have been reported[11], our analyses explicitly searched for sex-dependent signatures. Immunological resolution after infection may unfold over time even after symptoms subside, and there were indeed parameters that showed evidence of continued change in our cohort—defined as those that were correlated with time since COVID-19 diagnosis (TSD; Methods and Supplementary Table 1), including, as expected, SARS-CoV-2-neutralizing antibody titres[12] (Extended Data Fig. 1b). However, we were primarily interested in uncovering persistent, TSD-independent post-COVID-19 immune imprints, and we therefore focused on temporally stable immune states associated with previous mild COVID-19 but not correlated with TSD. Thus, we evaluated the differences between (1) female participants who had recovered from COVID-19 (COVR-F) versus healthy control female participants (HC-F); (2) male participants who had recovered from COVID-19 (COVR-M) versus healthy control male participants (HC-M); and (3) COVR-M versus COVR-F after accounting for male–female differences in healthy control individuals (hereafter, sex differences; Supplementary Table 2). The frequencies of myeloid cells such as monocytes and conventional/myeloid dendritic cells (cDCs) tended to be higher in the COVR-M group compared with the HC-M and/or COVR-F groups (Fig. 1c,d and Extended Data Fig. 1c,d), consistent with reports of myeloid cell disruption in COVID-19, particularly in severe, acute disease[13]. Here, a male-specific elevation in monocyte frequencies was detected even months after recovery from mild disease.

WBT data also revealed sex-dependent signatures associated with previous mild COVID-19 (Extended Data Fig. 1e; for example, the monocyte-related M11.0 and M4.0 from the blood transcriptional module (BTM) collection), including metabolic signatures such as oxidative phosphorylation (Supplementary Table 3). WBT differences can be driven by both cell composition and cell-intrinsic transcriptional changes. Indeed, the innate immune, metabolic and T cell-related signatures are driven, at least in part, by the increased circulating monocyte frequencies and correspondingly lower T cell frequencies in the COVR-M group (Fig. 1d and Extended Data Fig. 1f) because these transcriptional enrichment signals became statistically insignificant when monocyte frequencies were taken into account (data not shown).

To assess transcriptional alterations independent of cell frequencies, we used CITE-seq to examine the cell-type-specific contributions underlying the WBT signatures seen above. We clustered single cells and annotated the resulting clusters using surface protein expression profiles (Fig. 1e and Methods). Cell-type-specific transcriptional analysis pointed to both sex-dependent and -independent differences between participants who had recovered from COVID-19 and healthy control individuals (Supplementary Table 4). Among the enriched gene sets from the WBT analysis above (Extended Data Fig. 1e), but now free of cell-frequency confounding, the BTM M11.0/4.0 gene sets exhibit depressed expression in both classical and non-classical monocytes in participants who had recovered from COVID-19 relative to healthy control individuals in both sexes, whereas the converse is true for genes in the T cell activation signature (BTM M7.3) in both $CD8^+$ central memory and effector memory (EM) T cells (Fig. 1f–i, Extended Data Fig. 1g and Supplementary Table 5). The T cell activation signature in $CD8^+$ EMs was particularly pronounced in the COVR-M group (Fig. 1i). The genes driving the monocyte repression enrichment (that is, the leading-edge genes (LEGs)) include numerous surface receptors, such as those encoding pattern recognition receptors (*TLR2*, *TLR4* and *TLR8*), the peptidoglycan-recognizing receptor (*NOD2*), the high-affinity IgE FC receptor (*FCER1G*) and C-type lectin receptor (*CLEC4E*) (Fig. 1f,g). This innate immune receptor (IIR) signature in the monocytes, as well as the T cell activation signature, are predominantly not associated with TSD in both male and female individuals (Extended Data Fig. 1h).

The T cell activation signature probably emerged during and persisted after acute COVID-19 (ref. [14]), but this was less clear for the IIR signature. We therefore examined whether this signature could be linked to gene expression changes seen in acute COVID-19. Using a previously published CITE-seq dataset that we generated from an older, male-biased cohort of individuals from Italy with severe COVID-19 who were hospitalized[15], we noted that, within the classical monocytes, the average expression of the IIR LEGs from above was significantly lower in patients with acute COVID-19 than in healthy control individuals, and was negatively associated with disease severity (Extended Data Fig. 1i). Thus, this depressed IIR signature could have originated from and stably persisted since the acute response to the infection. Previous studies have reported several (potentially overlapping) types of altered monocytes in acute COVID-19, including those with lower antigen presentation, depressed NF-κB/inflammation or myeloid-derived suppressor-cell-like phenotypes[13,16,17]. However, none of these monocyte phenotypes was significantly different in the monocytes of participants who had recovered from COVID-19 compared with healthy control individuals in our cohort at the baseline before influenza vaccination (Supplementary Fig. 2), suggesting that our depressed monocyte gene signature involving pattern recognition and IIR genes is distinct from those identified earlier in acute disease. Together, our findings suggest that, even mild, non-hospitalized SARS-CoV-2 infections may establish new, temporally stable, sex-dependent immunological imprints detectable months after clinical recovery.

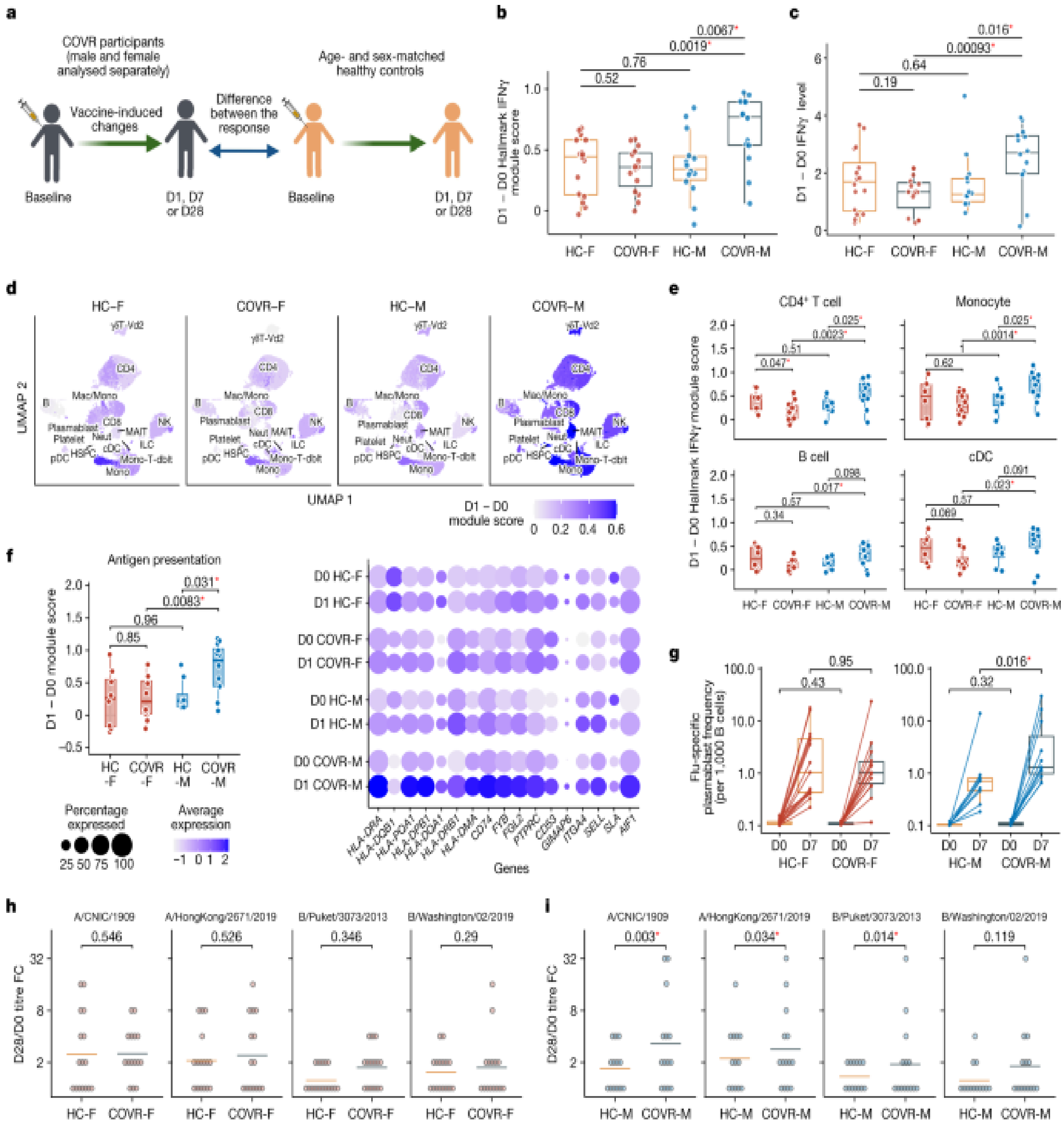
To assess whether other natural respiratory viral infections may leave similar unresolved sex-specific immune states, we used a published WBT dataset assessing two independent cohorts of patients with confirmed community influenza A (predominantly pandemic H1N1) infection during two different seasons[18] (2009–2010 and 2010–2011; Extended Data Fig. 2a). By comparing the WBT profiles before and after each season (that is, before infection and after recovery), we found robust post-infection changes consistent between these two independent cohorts in male individuals only (the changes in female individuals were not consistent between these two cohorts/seasons; Extended Data Fig. 2b and Supplementary Table 6). The genes with increased expression after recovery in male individuals were also enriched for genes that were more highly expressed in the COVR-M group

compared with the COVR-F group in our cohort (after accounting for the expected sex differences present in healthy participants; Extended Data Fig. 2c). Moreover, the genes with lower expression after recovery from influenza infection in males were enriched for the depressed IIR signature above, including *TLR5* and *VCAN* (Fig. 1f,g and Supplementary Table 6). These observations provide independent support that exposure to a respiratory viral pathogen can lead to persistent immunological imprints that are detectable in the blood, even in healthy individuals with mild disease. However, different viral infections, for example, those with distinct tropisms and inflammatory presentations, are also likely to leave pathogen-dependent imprints with distinct genes and processes; for example, the overlapping signals between post-influenza and post-mild COVID-19 are only a small subset of the sex-specific post-COVID-19 changes that we detected.

Contrasting influenza vaccination responses

We next examined whether previous COVID-19 may impact an individual's response to non-SARS-CoV-2 immunological challenges. The study participants received the seasonal influenza quadrivalent vaccine and were followed longitudinally for up to 100 days, including day 1 (D1), D7 and D28, to assess the vaccine response at the serological, molecular and cellular levels (Figs. 1a,b and 2a). This vaccine was selected in part due to its public health importance—the 2020–2021 influenza season was approaching at the start of our study and it was not clear whether previous COVID-19 infection would affect influenza vaccine responses. Moreover, the responses to seasonal influenza vaccination have been well characterized in healthy adults, including early innate/inflammatory and interferon (IFN) responses on D1 after vaccination and a strong but transient plasmablast peak around D7 culminating in the generation of influenza-specific antibodies[19,20]. Thus, influenza vaccination provides an excellent perturbation to probe the functional impacts of previous mild SARS-CoV-2 infection.

Fig. 2: Sex-specific response differences to influenza vaccination in individuals who had recovered from COVID-19 and matched control participants.



a, Schematic of the sex-specific comparisons of vaccine-induced changes from the baseline at timepoints after vaccination (D1, D7 and D28) between participants who had recovered from COVID-19 and healthy control participants. Analyses were applied to participants aged under 65 years (because older subjects received a higher dose vaccine; Methods). **b**, The D1 whole-blood IFNγ response transcriptional score (D1 − D0, computed using genes from the Hallmark IFNγ response gene set) for the COVR-F (*n* = 15), COVR-M (*n* = 14), HC-F (*n* = 16) and HC-M (*n* = 14) groups. **c**, The D1 response (D1 − D0) of serum IFNγ protein levels for the participants shown

in **b**. **d**, Surface-protein-expression-based UMAP analysis (as in Fig. 1e) with cells coloured according to the D1 IFNγ response transcriptional score (D1 – D0; see **b** for the gene set used) within each cell subset for the HC-F ($n = 8$), COVR-F ($n = 12$) and HC-M ($n = 8$) and COVR-M ($n = 12$) groups. Darker colour indicates a greater difference between D1 and D0 for the indicated cell subset. Mac, macrophages; Mono, monocytes. **e**, Similar to **b**, but for the indicated cell subsets (computed using the CITE-seq pseudobulk mRNA expression data for the cell subset) in the HC-F ($n = 8$), COVR-F ($n = 12$), HC-M ($n = 8$) and COVR-M ($n = 12$) groups. **f**, The D1 transcriptional response score (D1 – D0) of the antigen-presentation-related genes in classical monocytes for the same participants in **e** (left) (Methods). Right, the averaged expression of individual LEGs from the antigen-presentation genes (Methods) in classical monocytes. **g**, Influenza-specific plasmablast (PB; all $HA^{+}CD27^{+}CD38^{+}CD20^{low}CD21^{low}$ cells; Methods and Supplementary Fig. 3) frequencies at D0 and D7, plotted separately for the COVR-F ($n = 14$), HC-F ($n = 15$), COVR-M ($n = 11$) and HC-M ($n = 9$) groups. The lines connect data points from the same participant at D0 and D7. **h**, Analysis of the D28/D0 microneutralization titre fold change (FC) for each of the four strains in the seasonal influenza vaccine (columns) in the COVR-F and HC-F groups. Each dot represents one individual. The orange and grey lines indicate the average fold change for the HC-F and COVR-F groups, respectively. Unadjusted *P* values were derived from generalized linear models accounting for age, race, influenza vaccination history and baseline influenza titres (Methods). **i**, Similar to **h**, but for the COVR-M and HC-M groups. All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × IQR from box limits in each direction (upper and lower whiskers). Unadjusted *P* values are shown. Unless otherwise noted, the statistical significance of the difference between groups was determined using two-tailed Wilcoxon rank-sum tests. Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagram in **a** was created using BioRender.

WBT, peripheral immune cell frequency, CITE-seq, influenza-specific B cell and antibody titre analyses (assessing responses on D1, D7 and D28 relative to D0) together pointed to coordinated, sex-specific innate and adaptive response differences to the vaccine, with the COVR-M group generally mounting a more potent response compared with their healthy counterparts

and the COVR-F group (Fig. 2b–i, Extended Data Fig. 3a,c,d,g and Supplementary Tables 7 and 8). These include stronger innate/inflammatory and particularly IFN-related transcriptional responses (Fig. 2b and Extended Data Fig. 3a), with corresponding greater increases in circulating IFNγ protein levels in the serum by D1 in the COVR-M group (Fig. 2c). This systemic increase in IFNγ affects diverse cell types expressing the IFNγ signalling components as revealed by single-cell CITE-seq analysis—most peripheral immune cells had higher IFN response signatures on D1 in the COVR-M group compared with the other groups (based on comparing D1 versus D0; Fig. 2d; Fig. 2e shows $CD4^+$ T cells, B cells, monocytes and cDCs as examples). Baseline, prevaccination IFN-related transcriptional activity was largely indistinguishable between the participants who had recovered from COVID-19 and healthy control individuals (Extended Data Fig. 3b). Furthermore, a more robust response was observed for antigen-presentation genes, including both MHC class I and II genes in classical monocytes of the COVR-M group (Fig. 2f). Thus, individuals in the COVR-M group mount a stronger circulating IFNγ and corresponding transcriptional response in both innate and adaptive immune cells by D1 after influenza vaccination.

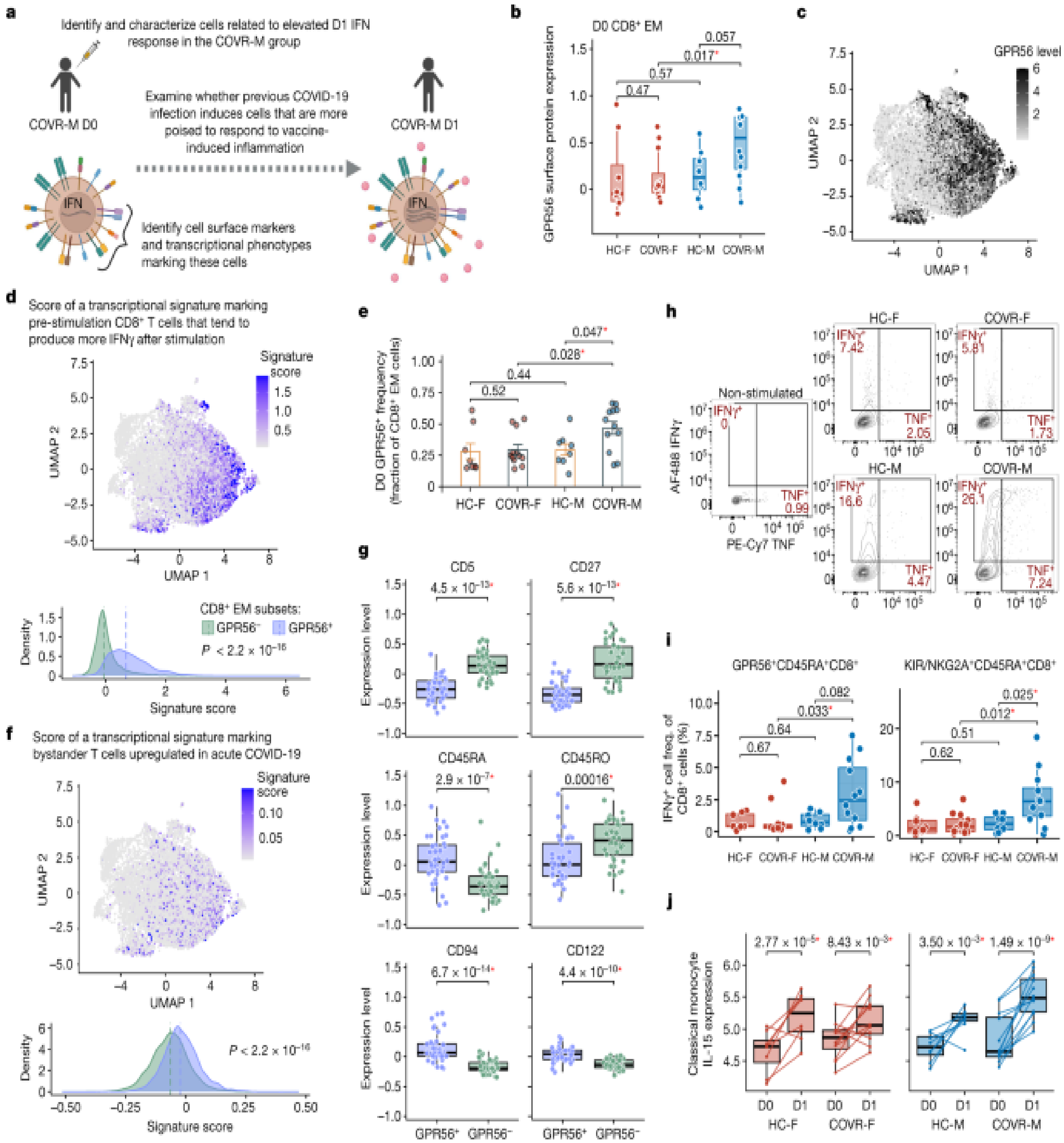
On the basis of previous studies of influenza vaccination in healthy adults and because heightened innate immune responses elicited by adjuvants are known to enhance adaptive responses[21], we hypothesized that the stronger early inflammatory responses in the COVR-M group would lead to a more robust humoral response. Indeed, we saw increased D7 B cell-related and plasma-cell-related transcriptional signatures in the COVR-M group (Extended Data Fig. 3a,c). Furthermore, the COVR-M group had a greater increase in influenza-specific plasmablasts compared with the HC-M group at D7 (Fig. 2g and Supplementary Fig. 3). Consistent with previous observations in healthy adults[22] and the hypothesis that the stronger early IFN response in the COVR-M group could help to induce a more robust B cell response, we detected a positive correlation between those two parameters, including the extent of influenza-specific plasmablast increases (Extended Data Fig. 3d). Consistently, the COVR-M group also had higher influenza-specific antibody responses compared with the HC-M group across all but one of the vaccine strains at D28 relative to the baseline (Fig. 2h,i, Methods, Extended Data Fig. 3e–g and Supplementary Table 8).

Although influenza infection and vaccination history can influence influenza vaccine responses[23], they alone are unlikely to explain the above findings because the COVID-19-recovered and healthy control groups had similar baseline antibody titres (Extended Data Fig. 3e,f), were age/sex-matched and were drawn from the same geographical region with very low influenza infection/transmission during the 2020–2021 season[8]. Moreover, the statistical model used to assess titre response differences incorporated prevaccination influenza titres as a covariate (Methods). The extent of time-dependent immune resolution after COVID-19 was probably not a factor because TSD and D28 titre responses are not correlated in either sex (data not shown). Together, these observations demonstrate that previous mild infection by SARS-CoV-2 can result in sex-dependent, coordinated changes in both innate and adaptive responses to immunization with non-SARS-CoV-2 antigens months after acute disease.

Linking the baseline to innate response

Having established that previous mild COVID-19 is associated with new baseline immune states before influenza vaccination (Fig. 1 and Extended Data Fig. 1) and COVR-M-group-specific responses after vaccination (Fig. 2 and Extended Data Fig. 3), we next attempted to link the two and examined what baseline variables and cellular circuits may contribute to the heightened IFN-related responses in the COVR-M group that could subsequently contribute to their more robust humoral responses (Fig. 3a). Using flow cytometry (Supplementary Fig. 1) and CITE-seq data, we first used a multivariate linear model to identify baseline/prevaccination immune cells of which the frequency predicted the D1 IFN-related responses (D1 versus D0 in serum IFNγ protein levels and IFN transcriptional signature score). A subset of $CD8^{+}$ T cells with an EM phenotype ($CD45RA^{-}CCR7^{-}CD28^{+}CD27^{-}$; early effector-like) was a top candidate in the COVR-M group and could therefore be a cellular source of IFNγ after vaccination (Extended Data Fig. 4a,b and Supplementary Fig. 4); the same relationship was not observed in the healthy control individuals (Supplementary Fig. 5a,b).

Fig. 3: Contributors to increased day 1 IFNγ responses in male participants who had recovered from COVID-19.



a, Schematic of the approach to assess why the COVR-M group had elevated early IFNγ responses. **b**, Comparison of the sample means of GPR56 surface expression in CD8^{+} EM T cells at D0 for the COVR-F (n = 12), HC-F (n = 8), COVR-M (n = 12) and HC-M (n = 8) groups. **c**, UMAP analysis of the D0 surface GPR56 protein expression on CD8^{+} EM cells from all 40 participants with CITE-seq data. The UMAP was derived using the top 60 variable surface proteins within the CD8^{+} EM cells (Methods). **d**, UMAP analysis as described in **c**, but showing the D0 gene-expression signature score computed using genes associated with CD29^{high}CD8^{+} T cells

identified earlier in an independent study[25] (Methods) (top). Density plot showing the distribution of the signature score above in the GPR56^{+}CD8^{+} and GPR56^{-}CD8^{+} EM cells (bottom). The dashed line indicates the median of the distribution. The statistical significance of the signature-score difference between the two cell subsets was determined at the single-cell level. **e**, Comparison of the proportion of GRP56^{+} cells (as fractions of CD8^{+} EM cells in the CITE-seq data) between the same participants as in **b** at D0. The error bars indicate the s.e.m. of each group. **f**, Similar to **d**, but showing the bystander T cell signature score at the baseline (D0) (signature genes originated from refs. [26,27]; Methods). **g**, Comparison of the average expression of the indicated memory cell-surface protein markers for the GPR56^{+}CD8^{+} versus GPR56^{-}CD8^{+} EM cells at D0 for the same participants as in **b**. Each point represents a participant. **h**, Representative flow cytometry contour plots of IFNγ^{+} and TNF^{+} gates within GPR56^{+}CD45RA^{+}CD8^{+} T cells after IL-15 stimulation in vitro in the indicated groups. The number shown for each gate denotes the percentage of parent cells (that is, GPR56^{+}CD45RA^{+}CD8^{+} T cells). **i**, The frequencies of IFNγ^{+}GPR56^{+}CD45RA^{+} VM-like CD8^{+} T cells (left; as fractions of CD8^{+} T cells) and IFNγ^{+}KIR/NKG2A^{+}CD45RA^{+}CD8^{+} T cells (right; as fractions of CD8^{+} T cells) in the same participants as in **b** after IL-15 stimulation in vitro. **j**, Comparison of D0 and D1 pseudobulk *IL15* mRNA expression (*y* axis) in classical monocytes for the same participants as in **b**. Significance was determined using a linear model accounting for age, race and influenza vaccination history (Methods). All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × IQR from box limits in each direction (upper and lower whiskers). Unless otherwise noted, the statistical significance of difference between groups was determined using two-tailed Wilcoxon rank-sum tests. Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagram in **a** was created using BioRender.

We next focused on all of the CD8^{+} T cells from clusters with an EM phenotype (CD8^{+} EM cells) in the CITE-seq data based on both surface protein markers and mRNA expression (Methods; the top cluster protein markers are shown in Supplementary Table 10). We searched for differences in average surface marker expression of cells in these CD8^{+} EM clusters

across the four participant groups and found that GPR56 was the top differentially expressed marker with increased expression in the COVR-M group relative to the HC-M and COVR-F groups (Fig. 3b,c and Supplementary Table 10). This was intriguing because $CD4^+$ EM and TEMRA (terminally differentiated EM cells re-expressing CD45RA) T cells marked by surface GPR56 expression at the baseline (before stimulation) have been reported to produce increased amounts of IFNγ after stimulation with PMA and ionomycin (PMAI)[24]. Consistent with this, $GPR56^+CD8^+$ EM cells in our data are enriched for a transcriptional signature (derived in an independent study[25]) that marks $CD8^+$ EM cells poised to secrete higher levels of IFNγ after PMAI stimulation (Fig. 3d). Thus, $GPR56^+CD8^+$ EM cells could be a source of elevated IFNγ production in the COVR-M group after influenza vaccination. Indeed, the frequency of these cells was elevated in the COVR-M group relative to in the HC-M and COVR-F groups before vaccination (Fig. 3e), but was not correlated with the TSD and was therefore temporally stable (assessed by Spearman's correlation: $P = 0.18$ (COVR-F) and $P = 0.51$ (COVR-M)). Moreover, *IFNG* transcripts increased significantly in these cells on D1 after influenza vaccination in the COVR-M group (Extended Data Fig. 4c,d). These data suggest that previous COVID-19 increases the frequency of $GPR56^+CD8^+$ EM cells in male individuals and these cells are poised to make more IFNγ early after influenza vaccination, which together contributed to the higher IFNγ production in the COVR-M group; consistent with this hypothesis, this was not observed in $GPR56^-$ cells (Extended Data Fig. 4d and Supplementary Fig. 5c).

Mild, non-hospitalized COVID-19 has been reported to induce bystander activation (non-SARS-CoV-2 specific) of $CD8^+$ T cells[26]. Notably, the $GPR56^+$ cells are also enriched for a transcriptional signature associated with bystander T cell activation[26,27] (Fig. 3f). Moreover, $GPR56^+CD8^+$ EM cell frequency is positively correlated with the T cell activation signature score, which was elevated at the baseline in the COVR-M group as shown above (Fig. 1i and Extended Data Fig. 4e). This suggests that some of these cells may have expanded in a bystander manner during the acute phase of the infection. This prompted us to consider whether these $GPR56^+$ cells are similar to bystander-activated virtual memory (VM) $CD8^+$ T cells, a feature of which is their ability to be activated rapidly by inflammatory cytokines alone (for example, IL-12, IL-18 and IL-15) to produce IFNγ without T cell

receptor (TCR) stimulation[28,29]. VM CD8$^+$ T cells expand through cytokine stimulation, including IL-15 induced by viral infection (IL-15 concentrations are known to be elevated in patients with acute COVID-19 and are correlated with disease severity[30]), and are characterized by a differentiated EM phenotype expressing CD45RA[28]. We assessed several reported surface markers of these cells[28] in GPR56$^+$ versus GPR56$^-$ cells and found that the GPR56$^+$ cells were indeed phenotypically similar to VM cells (Fig. 3g). For example, GPR56$^+$ cells have higher CD122 but lower CD5 surface expression compared with their GPR56$^-$ counterparts; CD5 surface expression has been linked to the extent of previous IL-15 (or potentially other inflammatory cytokine) encounters[28,31]. Notably, on the basis of the surface levels of CD45RA and CD45RO, the GPR56$^+$ cells appear to situate phenotypically between GPR56$^-$ and TEMRA cells (Extended Data Fig. 4f).

To further test our hypothesis, we performed in vitro stimulation experiments to assess whether GPR56$^+$CD8$^+$ T cells can produce IFNγ in response to several cytokines that are known to be induced by vaccination or infection (Supplementary Fig. 6a). Stimulation with IL-15 showed that GPR56$^+$CD45RA$^+$CD8$^+$ T cells from the COVR-M group produced more IFNγ compared with those from the COVR-F group (Fig. 3h,i). CD8$^+$ VM-like T cells were identified using the surface markers CD45RA$^+$, KIR$^+$ and/or NKG2A$^+$[32,33] and the COVR-M group produced higher levels of IFNγ in these cells (Fig. 3i). Stimulation with IL-12, IL-15 and IL-18 together showed similar trends (Supplementary Fig. 6b). Stimulation with IL-18 alone or IL-12 and IL-18 together also showed similar trends, but these conditions induced less robust IFNγ than IL-15 stimulation (data not shown). We next assessed the cellular source of IL-15 post-influenza vaccination using CITE-seq data and found that classical monocytes from the COVR-M group showed the most significant increases in *IL15* mRNA levels on day 1 after influenza vaccination (Fig. 3j). Together, this suggests that the increased IFNγ response in the COVR-M group after vaccination could be attributed to increased baseline (prevaccination) frequencies of cells that are also intrinsically more responsive to inflammatory stimulation, including classical monocytes that produce elevated IL-15 and CD8$^+$ VM-like T cells that mount a more robust IFNγ response to cytokine stimulation alone.

As VM T cells can be rapidly activated to produce cytokines without clonal, antigen-specific expansion[28], we assessed the clonality of the $GPR56^{+}CD8^{+}$ EM cells at different timepoints after influenza vaccination using V(D)J/TCR data from CITE-seq. The clonality of both the $GPR56^{+}CD8^{+}$ EM and TEMRA cells remained stable across D0 (before vaccination), D1 and D28 after influenza vaccination (Extended Data Fig. 4g,h). The frequencies of $GPR56^{+}CD8^{+}$ EM clones shared across timepoints within individuals were also similar (Extended Data Fig. 4i). Together, these data argue against the notion that the heightened activation of the $GPR56^{+}$ cells early after influenza vaccination in the COVR-M group was due solely to TCR-dependent T cell activation and clonal expansion. As was shown previously[28,29] and above in our in vitro stimulation data, a more plausible explanation is that these $CD8^{+}$ VM-like cells were activated to produce IFNγ by the inflammatory cytokines elicited by the influenza vaccine in an antigen-independent manner. Despite their resemblance to VM cells, some of the $GPR56^{+}$ cells could have developed from naive cells through conventional, non-bystander pathways (for example, some could be developed during acute COVID-19 and are specific for SARS-CoV-2), although none of these cells had a CDR3 sequence that matches a public clone deemed to be specific for SARS-CoV-2 (data not shown). Bona fide, antigen-specific memory $CD8^{+}$ T cells developed from naive cells through TCR stimulation have also been shown to produce IFNγ in response to inflammatory cytokines alone in mice[34,35].

Our data also revealed other cell types that could have contributed to the increased IFNγ production observed on D1 after vaccination in the COVR-M group (Supplementary Fig. 7a–c). *IFNG* transcripts increased more in the COVR-M group compared with in the HC-M and COVR-F groups on D1 in $CD16^{low}$ natural killer (NK) cells (Supplementary Fig. 7c and Supplementary Table 4). Moreover, the baseline frequency of $CD16^{low}$ NK cells was correlated with the extent of the D1 increase in both *IFNG* expression and serum protein levels (Supplementary Fig. 7b). However, the IFNγ response in total NK cells after IL-15 stimulation in vitro was not significantly higher in the COVR-M group (Extended Data Fig. 4j), probably because $CD16^{low}$ NK cells are a small subset of total NK cells. By contrast, IL-15 stimulation in vitro revealed a higher IFNγ response in MAIT cells in the COVR-M group compared with the COVR-F and HC-M

groups (Extended Data Fig. 4j), but the increase in *IFNG* mRNA expression on D1 after influenza vaccination was not statistically significant in the COVR-M group based on CITE-seq data (Supplementary Fig. 7c). $CD8^+$ T cells with a TEMRA ($CD45RA^+CD45RO^-CCR7^-$) phenotype might also have a role as their IFNγ response after IL-15 stimulation in vitro was higher in the COVR-M group compared with in the COVR-F and HC-M groups (Extended Data Fig. 4j), which is consistent with the CITE-seq data (Supplementary Fig. 7c).

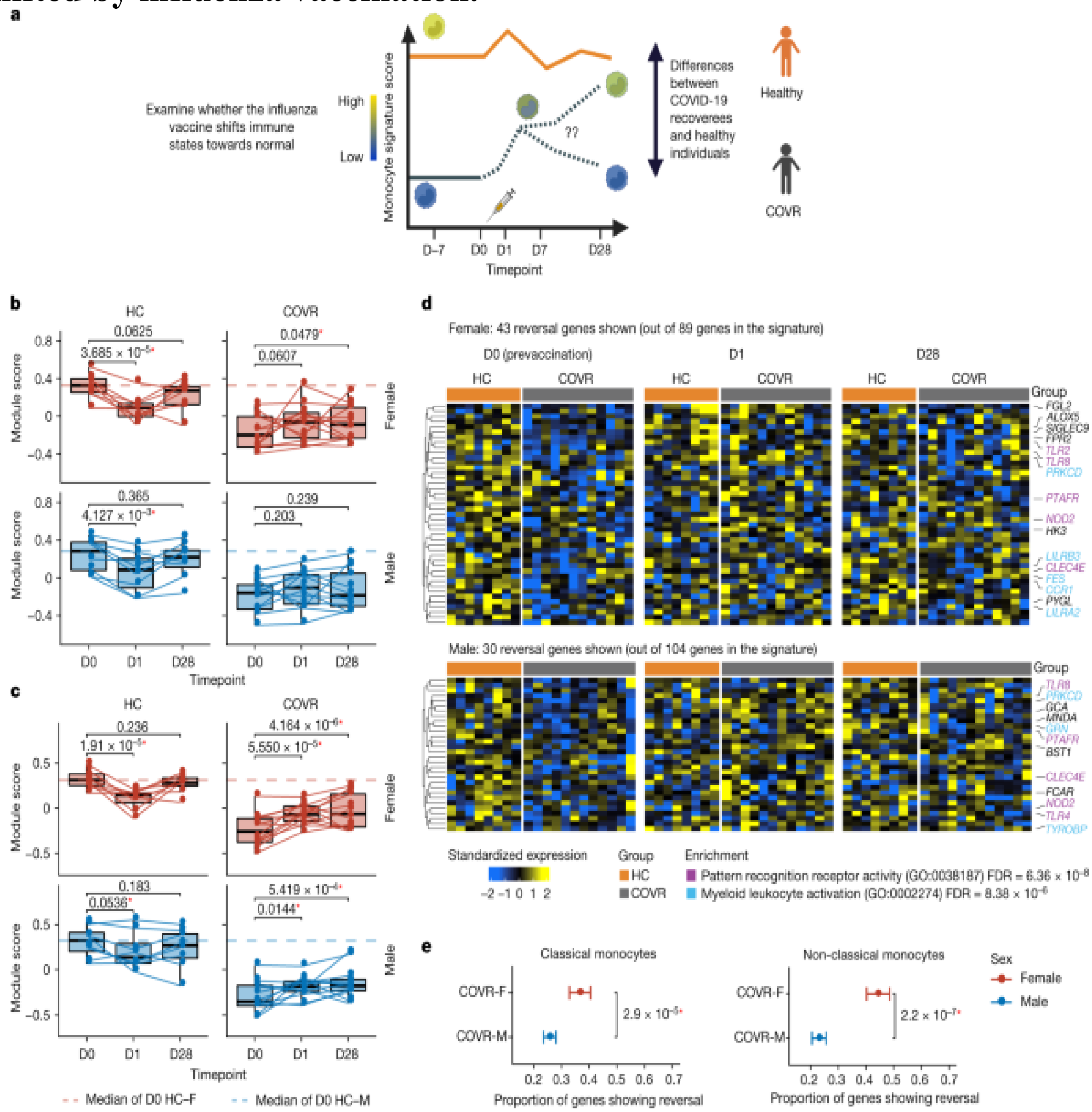
Taken together, we demonstrate a population of $CD8^+$ EM T cells marked by GPR56 expression and VM-like markers with antigen-agnostic pro-inflammatory potential after heterologous vaccination. Importantly, these cells, and potentially $CD16^{low}$ NK, MAIT and $CD8^+$ TEMRA cells (albeit with less support from our CITE-seq data), emerged in otherwise clinically healthy individuals and are especially elevated and more poised to respond in male individuals who were months recovered from mild SARS-CoV-2 infection, providing additional evidence for sex-specific, functionally relevant immune set points linked to previous mild COVID-19.

Vaccination shifts monocyte imprints

Given the potential for vaccine-induced training effects[6,36,37], we next examined whether influenza vaccination can alter some of the post-COVID-19 transcriptional imprints that we detected earlier (Fig. 4a). We focused on the monocytes owing to the robustly depressed IIR signature reported above (in participants who had recovered from COVID-19 versus healthy control participants; Fig. 1f,g) and because vaccines can potentially induce long-lasting changes in these cells[6,36]. Using the healthy control baseline (D0) as a healthy reference, we used CITE-seq data to assess the average expression of the signature genes (identified above) before and after vaccination in participants who had recovered from COVID-19, separately for classical (Fig. 1f) and non-classical monocytes (Fig. 1g) in male and female individuals (Extended Data Fig. 5a,b). As was observed above, these genes had lower average expression in participants who had recovered from COVID-19 compared with healthy control individuals in both sexes at D0 before vaccination. However, their average expression increased towards that of the healthy control individuals by D1 and persisted until D28 in the

COVR-F and COVR-M groups, although the effect appeared to be stronger in the COVR-F group (Extended Data Fig. 5a,b).

Fig. 4: Post-mild-COVID-19 gene expression imprints in monocytes shifted by influenza vaccination.



a, Schematic of the study questions. **b**, The module scores of the IIR signature (Fig. 1f) in the HC-F ($n = 8$), HC-M ($n = 8$), COVR-F ($n = 12$) and COVR-M ($n = 12$) groups at D0, D1 and D28 using the CITE-seq pseudobulk gene expression data in classical monocytes. The dashed line represents the median D0 score of the healthy control individuals of the

same sex. The lines connect data points from the same participant at different timepoints. The statistical significance of differences was determined using a mixed-effects model accounting for age, race and influenza vaccination history (Methods). Unadjusted *P* values are shown. **c**, Similar to **b**, but for non-classical monocytes (Fig. 1g). **d**, Heat map showing the expression of the reversal genes in classical monocytes (row-standardized; see Extended Data Fig. 5c for non-classical monocytes). Reversal genes are defined as those genes in the baseline IIR signature (Fig. 1f) of which the expression in participants who had recovered from COVID-19 at D1 and D28 after vaccination moved towards the baseline (prevaccination) expression of healthy control individuals. The COVR-F (top) and COVR-M (bottom) groups are shown separately; healthy control individuals are also included for comparison. The rows are genes and columns are individual samples (grouped by participant/timepoint) with timepoint and participant group labels shown at the top, including the same participants as in **b** at each timepoint. The names of genes that belong to gene sets of functional interest are shown. False-discovery-rate-corrected enrichment *P* values are shown. **e**, Comparison of the proportion of IIR signature genes (Fig. 1f,g) that show partial reversal in the COVR-F versus COVR-M groups in classical and non-classical monocytes. The mean and 95% confidence intervals (denoted by the bars) were derived from a bootstrapping procedure (Methods). Significance was determined using two-tailed Wilcoxon tests between the bootstrapped samples. All of the box plots show the median (centre line), first and third quantiles (box limits), and max 1.5 × IQR from box limits in each direction (upper and lower whiskers). Significant ($P < 0.05$) differences are highlighted with a red asterisk (*). The diagram in **a** was created using BioRender.

Quantifying the average expression (module score) of these sex- and cell-type-dependent gene sets (Fig. 1f,g) within individual participants over time confirmed a similar and significant trend of shift towards the healthy control individuals (Fig. 4b,c). This analysis further revealed that the extent of this change in gene expression was more pronounced in the non-classical than in the classical monocytes (Fig. 4b,c). Notably, the behaviour of these genes was divergent in the healthy control individuals—the gene module score trended lower on D1 and reverted to prevaccination levels by day 28 in the healthy control individuals (Fig. 4b,c). Although the underlying mechanism of this divergence is unclear, the monocytes in healthy control individuals

could have responded to the vaccine-induced inflammation by downregulating certain immune receptor genes and associated signalling genes in a negative feedforward mechanism to avoid over-responding, while the 'depressed' monocytes in participants who had recovered from COVID-19 instead responded by increasing the expression of these genes and therefore moving towards the normal (healthy baseline) level.

We next identified the individual genes within these gene sets that moved towards the healthy control baseline (Methods). In both classical and non-classical monocytes, the fraction of reverting genes was significantly higher in female compared with male participants (Fig. 4d,e and Extended Data Fig. 5c), although several TLRs (for example, *TLR2* and *TLR4*) and *NOD2* were significant in both sexes in one or both monocyte subsets. These changes were probably not due to continued immune resolution after infection because the baseline (D0) expression of these genes did not correlate with TSD (Extended Data Fig. 1h), and they increased acutely by D1 after vaccination and persisted to D28. Notably, in contrast to this depressed IIR signature (Fig. 1f,g and Extended Data Fig. 1i), other monocyte-related transcriptional signatures that are known to have lower expression during acute COVID-19—such as genes related to antigen presentation, inflammatory and NF-κB activation and myeloid suppressor cells[13,15,16,17,38,39]—were similar between participants who had recovered from COVID-19 and healthy control individuals at D0/baseline; vaccination also did not consistently elicit longer-lasting changes in these signatures out to D28, although the COVR-M group tended to have elevated antigen presentation transcriptional responses in non-classical monocytes on D1 that remained mildly elevated by D28 (Extended Data Fig. 5d,e).

Together, CITE-seq analysis revealed that the early (D1) response to influenza vaccination elevates a set of previously (that is, before vaccination) depressed IIR genes in the monocytes of participants who had recovered from COVID-19 out to at least D28 after vaccination. Although the functional relevance of these changes remains to be determined, these results suggest that the early inflammatory responses to influenza vaccination can help to shift the post-COVID-19 immune state of monocytes towards that of healthy individuals, particularly in female recoverees.

Discussion

Although both acute and longer-term immune perturbations in hospitalized patients with COVID-19 have been reported[13,40,41,42,43], less is known regarding healthy recovered individuals with previous mild, non-hospitalized SARS-CoV-2 infection months after acute illness, without confounding comorbidities such as obesity, autoimmunity or immunodeficiency. Here we reveal that clinically healthy recoverees of previous non-hospitalized COVID-19 possess sex-specific immune imprints beyond SARS-CoV-2-specific immunity, some of which become apparent only after vaccination with antigens that are distinct from SARS-CoV-2. Our findings are consistent with the sex dimorphic nature of acute responses to SARS-CoV-2 and other immune challenges[11]. Healthy female individuals tend to mount heightened inflammatory responses to infections and vaccines[44]; it was therefore surprising to find the qualitative opposite here in which the COVR-M was found to have a more poised immune status at the baseline and stronger innate and adaptive responses to influenza vaccination. Although persistent immune state changes (over months) in patients with long COVID have been reported[41], most of the individuals in our study reported no or minor post-COVID-19 sequelae. Future research could assess whether some of the sex-specific imprints, including differences in vaccination responses, are associated with long COVID[7].

Our findings suggest that the poised baseline immune states in the COVR-M group helped to establish the more robust IFN, plasmablast and antibody responses on days 1, 7 and 28, respectively, after influenza vaccination. The early IFN responses may be attributed to monocytes with higher *IL15* transcriptional responses early after vaccination coupled with elevated prevaccination frequencies of VM-like $CD8^+$ T cells poised to produce more IFNγ after IL-15 stimulation. The monocyte imprint that we described involving poised *IL15* mRNA production in male recoverees and the transcriptionally depressed innate receptor gene signature in both sexes are consistent with the notion of trained innate immunity[6]. Notably, although the latter signature could be detected in patients with acute COVID-19 with severe disease, it is distinct from the depressed antigen presentation or myeloid-suppressor-cell-like states found in previous studies of acute

COVID-19 (refs. [13,15,16,17,38,39]). As trained innate immunity can be mediated through myriad mechanisms, including chromatin and metabolic changes within cells, future studies could explore these potential mechanisms in monocytes, including the influences of sex/gender, acute disease severity and age among participants with a range of post-COVID-19 clinical sequelae. Given that the half-life of circulating monocytes is relatively short (and can be shorter than 28 days)[45], the partial reversal that we detected is possibly attributable to bone marrow myeloid progenitor cells, as haematopoietic stem and progenitor cells have been shown to exhibit chromatin accessibility changes after SARS-CoV-2 infection[46].

Bystander T cell activation has been reported after natural viral infections[47], including SARS-CoV-2 (ref. [26]). More recently, bystander-activated $CD8^+$ EM T cells have been identified to have an important role in controlling early infection, including VM cells that have no previous antigen exposure or TCR engagement[28,29]. As these cells can emerge after cytokine stimulation alone, it is possible that a stronger or more prolonged cytokine response to SARS-CoV-2 in male relative to female individuals during acute disease may have resulted in the elevated frequencies of the $GPR56^+CD8^+$ VM-like cells in the COVR-M group. This hypothesis is consistent with reports that male individuals hospitalized with COVID-19 tend to experience greater innate immune activation (as measured by circulating cytokines) compared with female individuals[48,49].

Some of the immune imprints that we observed could be shared among different types of viral infections, but some are probably unique to SARS-CoV-2, as suggested by our comparison with natural influenza infection. Our findings point to the possibility that any infection or immune challenge may change the immune status to establish new baseline set points encoded by the states of not only a single cell lineage, but also a network of interacting cell types such as VM T cells and monocytes. Moreover, although baseline immune statuses that are predictive of future responses are often different across and temporally stable within individuals over a timescale of months[50,51], our results suggest that such baseline immune states could have been established by past infections and are stable up to the next perturbation. Thus, the baseline immune status of an individual, with the potential to impact future responses in both antigen-specific and antigen-agnostic ways,

is shaped by a multitude of previous exposures[2,3]. In addition to revealing underlying principles regarding what happens after two well-defined natural immunological encounters—mild COVID-19 and influenza vaccination in humans—our observations provide a basis for studying more complex scenarios, such as what happens over longer timescales with additional inflammatory encounters. Our research brings forth the concept that even mild viral infections could establish new immunological set-points impacting future immune responses in an antigen-agnostic manner and illustrates how heterologous vaccination could be used as a tool to reveal such functional imprints.

Methods

Patient population and sample collection

Participants aged at least 18 years were recruited between August and December 2020 from the local area (Maryland, Virginia and the District of Columbia) and enrolled on National Institutes of Health (NIH) protocol 19-I-0126 (Systems analyses of the immune response to the seasonal influenza vaccine). The study was approved by the NIH Institutional Review Board (ClinicalTrials.gov: NCT04025580) and complied with all relevant ethical regulations. Informed consent was obtained from all of the participants. After informed consent was obtained, a baseline history and physical examination were performed. The participants were asked to characterize any present, persistent symptoms of past SARS-CoV-2 infection. Exclusion criteria included obesity (BMI ≥ 30); history of or suspicion of any autoimmune, autoinflammatory or immunodeficiency disease; history of any vaccine within the past 30 days (live attenuated) or 14 days (non-live attenuated); history of any experimental vaccine; history of a parasitic, amoebic, fungal or mycobacterial infection in the past year; or current infection. The COVID-19 vaccine was not available at the time of the study, and no study participants participated in any COVID-19 vaccine trials. All study visits occurred at the NIH Clinical Center (CC) in Bethesda, MD, USA. Blood samples were collected by phlebotomy staff at the NIH CC. The samples were collected between September 2020 and April 2021. No sample size calculations were done prior to enrolment, in part because there were no reliable effect size estimates related to the impact of prior COVID-

19 infection on vaccine responses. The number of subjects in the study was the number that were able to be recruited during the recruitment period.No blinding or randomization was performed.

Samples were collected from participants from three groups: (1) those with a previous history of symptomatic SARS-CoV-2 infection (defined as a history of a positive nasal PCR test and positive Food and Drug Administration (FDA) Emergency Use Authorization (EUA) SARS-CoV-2 antibody test at the time of protocol screening); (2) those with a history of asymptomatic SARS-CoV-2 infection (defined as testing positive using the FDA EUA SARS-CoV-2 antibody test at the time of the protocol exam, but with no history of COVID-like symptoms; no time since COVID-19 infection or diagnosis was identifiable for this group and they were excluded from all TSD analyses); and (3) individuals with no history of SARS-CoV-2 infection (defined as testing negative with the FDA EUA SARS-CoV-2 antibody test at the time of the protocol screening).

Blood for peripheral blood mononuclear cells (PBMCs), serum, whole-blood RNA (Tempus Blood RNA Tube, Thermo Fisher Scientific), complete blood count with differential (CBC) and lymphocyte phenotyping was collected at each of the following timepoints relative to seasonal influenza vaccination (day 0): days −7, 0, 1, 7, 14, 28, 70 and 100. Optional stool samples were collected at days 0, 28 and 100. The participants were provided with Cardinal Health Stool Collection kits (Cardinal Health) and Styrofoam storage containers with ice packs to collect stool samples at home and return in person to the NIH. After day 100, the participants had the option to continue to provide monthly blood samples for PBMCs, serum, whole blood RNA, CBC with differential and lymphocyte phenotyping through August 2021.

At each timepoint after study enrolment, data were collected and managed using the REDCap (v.8.5.27) electronic data capture tools hosted at the NIH[54,55]. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for data integration and

interoperability with external sources. REDCap electronic questionnaires were used to collect information from the participants through two separate IRB-approved surveys. A survey to evaluate vaccine-related adverse events or symptoms was administered on study days 1 and 7 and a separate survey to evaluate for any health changes or new medications was administered at every visit starting on day 0. Surveys were sent by email to the participants and the responses were transferred from the REDCap system to the NIH Clinical Research Information Management System (CRIMSON) system by the study team.

Influenza vaccination

Participants aged between 18 and 64 years were administered the Flucelvax Quadrivalent seasonal influenza vaccine (2020–2021; Seqirus). Participants aged 65 years and older were administered the high-dose Fluzone Quadrivalent seasonal influenza vaccine (2020–2021; Sanofi Pasteur).

Influenza microneutralization titres

Virus-neutralizing titres of pre- and post-vaccination sera were determined in a microneutralization assay based on the methods of the pandemic influenza reference laboratories of the Centers for Disease Control and Prevention (CDC) using low-pathogenicity vaccine viruses and MDCK cells. The X-179A virus is a 5:3 reassortant vaccine containing the *HA*, *NA* and *PB1* genes from A/California/07/2009 (H1N1pdm09) and the five other genes from A/PR/8/34 were donated by the high-growth virus NYMC X-157. Immune sera were also tested for neutralization titres of the seasonal vaccine strains H1N1 A/Brisbane/59/07, H3N2 A/Uruguay/716/07 and B/Brisbane/60/2001. Internal controls in all of the assays were sheep sera generated against the corresponding strains at the Center for Biologics Evaluation and Research, FDA. All individual sera were serially diluted (twofold dilutions starting at 1:10) and were assayed against 100 median tissue culture infectious dose of each strain in duplicates in 96-well plates (1:1 mixtures). The titres represent the highest dilution that completely suppressed virus replication.

SARS-CoV-2 pseudovirus production and neutralization assay

Human codon-optimized cDNA encoding SARS-CoV-2 S glycoprotein (GenBank: NC_045512) was cloned into eukaryotic cell expression vector pcDNA 3.1 between the BamHI and XhoI sites. Pseudovirions were produced by co-transfection of Lenti-X 293T cells with psPAX2(gag/pol), pTrip-luc lentiviral vector and pcDNA 3.1 SARS-CoV-2-spike-deltaC19, using Lipofectamine 3000. The supernatants were collected at 48 h after transfection and filtered through 0.45 µm membranes and titrated using 293T-ACE2 cells (HEK293T cells that express ACE2 protein). The following reagent was obtained through BEI Resources, NIAID, NIH: human embryonic kidney cells (HEK293T) expressing human angiotensin-converting enzyme 2, HEK293T-hACE2 cell line, NR-52511.

For the neutralization assay, 50 µl of SARS-CoV-2 S pseudovirions were pre-incubated with an equal volume of varying dilutions of serum at room temperature for 1 h, then virus–antibody mixtures were added to 293T-ACE2 cells in a 96-well plate. After incubation for 3 h, the inoculum was replaced with fresh medium. After 24 h, cells were lysed and luciferase activity was measured as previously described[56,57,58]. Controls included a cell-only control, virus without any antibody control and positive control sera. Lenti-X 293T cells were obtained from Takara Bio (Cat. No. 632180). 293T-ACE2 cells were obtained from ATCC. The 293T-ACE2 cells were checked for expression of ACE2 and validated by FACS analysis. Neither of the cell lines was authenticated by karyotyping or other genomic techniques. Both cell lines tested negative for *Mycoplasma*.

SPR-based antibody binding kinetics of human serum

Steady-state equilibrium binding of serum was monitored at 25 °C using the ProteOn Surface Plasmon Resonance (BioRad) system as previously described[59,60,61]. The purified recombinant SARS-CoV-2 or other proteins were captured to a Ni-NTA sensor chip (BioRad, 176-5031) with 200 resonance units (RU) in the test flow channels. The protein density on the chip was optimized such as to measure monovalent interactions independent of the antibody isotype. Serial dilutions (10-, 30- and 90-fold) of freshly prepared sample in BSA-PBST buffer (PBS pH 7.4 buffer with Tween-20 and BSA) were injected at a flow rate of 50 µl min^{-1} (120 s contact duration) for association, and disassociation was performed over a 600 s interval.

Responses from the protein surface were corrected for the response from a mock surface and for responses from a buffer-only injection. Total antibody binding was calculated using the BioRad ProteOn manager software (v.3.1). All SPR experiments were performed twice, and the researchers performing the assay were blinded to sample identity. Under these optimized SPR conditions, the variation for each sample in duplicate SPR runs was <5%. The maximum resonance units (max RU) data shown in the figures were the RU signal for the tenfold-diluted serum sample.

PBMC isolation

PBMC samples were isolated from blood collected in Vacutainer EDTA tubes (generic laboratory supplier) using SepMate-50 tubes (STEMCELL Technologies) with the following modifications to the manufacturer's protocol: the blood samples were diluted 1:1 with room temperate PBS and mixed by pipetting. The diluted blood was layered on top of a 15 ml Cytiva Ficoll PAQUE-Plus (Cytiva Life Sciences) layer in the SepMate tube. The SepMate tubes were centrifuged at 1,200*g* for 10 min with brake set to 5 at room temperature. After centrifuging, the top plasma layer was removed as much as possible without disturbing the PBMC layer. If there were any cells stuck on the wall of the tube, they were gently scraped from the wall using a pipette so they could be resuspended with the rest of the cells. The cells were poured from the SepMate tube in to a 50 ml conical tube. The tubes containing cells were filled up to 50 ml with cold wash buffer (PBS with 2% FBS) and mixed by inverting. The tubes were centrifuged at 300*g* for 10 min with brake set to 5 at room temperature. After centrifuging, the supernatant was removed without disturbing the cell pellet. After resuspending the pellet with cold wash buffer, the cells were counted using the Guava Muse Cell Analyzer (Luminex). The tubes were again centrifuged at 300*g* for 10 min with brake set to 5 at room temperature. The supernatant was removed without disturbing the cell pellet.

On the basis of the cell count, 6–10 million PBMCs were frozen per vial for each sample. As the cells were counted before the last centrifuging, a 50% cell loss was assumed and accounted for in the calculations from cell count. The cell pellet was resuspended with $n \times 600$ µl (where n is the number of PBMC vials to be frozen) freezing medium (RPMI with 10% FBS) by gentle

pipetting. After freezing the medium, $n \times 600$ µl DMSO freeze (FBS with 15% DMSO) was added drop-by-drop while gently shaking the tube. In other words, for each vial of PBMC that was to be frozen, 600 µl of freezing medium and 600 µl of DMSO freeze was added, bringing the total volume for each vial to 1.2 ml. The solution was gently mixed by pipetting before transferring 1.2 ml cell solution to each 1.8 ml cryovial (general laboratory supplier). The cell vials were placed into CoolCell Containers (Thomas Scientific) and the container was placed into a −80 °C freezer. After at least 4 h, the PBMC vials were transferred to liquid nitrogen.

RNA isolation

Blood was drawn directly into the Tempus Blood RNA Tube (Thermo Fisher Scientific) according to the manufacturer's protocol. Two Tempus tubes were collected at each study timepoint. The blood sample from each Tempus tube was aliquoted into two 4.5 ml cryovials (general laboratory supplier). These cryovials were directly stored at −80 °C.

The RNA samples were isolated in groups of 12–22 samples per batch based on careful batching before isolation to reduce confounding factors due to age, gender and patient group.

RNA was isolated from blood in the Tempus tube using the QIAsymphony RNA Kit (Qiagen) using the QIAsymphony SP instrument (Qiagen). Blood samples were thawed on ice before each sample was transferred to a 50 ml conical tube. The total volume of the sample was brought to 12 ml by adding 1× PBS. The tubes were vortexed at full speed for 30 s, followed by centrifugation at 3,500*g* for 1 h at 4 °C. After centrifugation, the supernatant from the tubes was decanted and the tubes were placed upside-down on clean paper towels for 2 min to allow residual liquid to drain. To resuspend the pellet, 800 µl of RLT+ buffer was added to the bottom of each tube and vortexed for few seconds. All 800 µl of each sample was transferred to 2 ml screw cap tubes (Sarstedt). The tubes were placed into #3b adapters (Qiagen) to be loaded onto the QIAsymphony system.

On the QIAsymphony system, the RNA CT 800 protocol was selected and used for RNA isolation. The instrument was set up according to the manufacturer's protocol and the elution volume for RNA samples was set to

100 µl. The final volume of the eluted RNA samples ranged from 65 µl to 95 µl.

RNA yields were determined using the Qubit RNA BR kit or Qubit RNA HS kit (Thermo Fisher Scientific) on the basis of the yield. RNA RIN numbers were measured using RNA ScreenTape (Agilent Technologies). The average RIN was 8.3 and the average yield was 81.3 ng μl^{-1} for the RNA samples.

RNA-seq

RNA-seq libraries were prepared manually using Universal Plus mRNA-Seq with NuQuant, Human Globin AnyDeplete (Tecan Genomics) according to the manufacturer's protocol. For each sample, 500 ng of total RNA was used to isolate mRNA by poly(A) selection. Captured mRNA was washed, fragmented and primed with the mix of random and oligo(dT) primers. After cDNA synthesis, ends were repaired and ligated with unique dual index adaptor pairs. Unwanted abundant transcripts from rRNA, mtRNA and globin were removed using AnyDeplete module. The remaining library was amplified by 14 cycles of PCR and purified with AMPure XP reagent (Beckman Coulter).

Library concentration was determined using the Quant-iT PicoGreen dsDNA Assay kit (Thermo Fisher Scientific) on the BioTek Synergy H1 plate reader (BioTek Instruments) using 2 µl sample. Library size distribution was determined using D1000 ScreenTape (Agilent Technologies) on the 4200 TapeStation System (Agilent Technologies). A total of 32 samples were randomly selected from each plate to measure the library size distribution. To determine fragment size, the region on the electropherogram was set from 200 bp to 700 bp. An average of the fragment sizes was used for the rest of libraries to calculate the molarity.

To create a balanced pool for sequencing, all of the libraries from one plate were diluted to the same molar concentration using the QIAgility liquid handling robot (Qiagen), and equal volumes of normalized samples were pooled. A total of 96 samples were pooled from each plate on plates 1–4 and 35 samples were pooled from plate 5. For an accurate quantification of the

pooled libraries, quantitative PCR was performed using the KAPA Library Quantification Kit (Roche).

All of the libraries were sequenced on the NovaSeq 6000 instrument (Illumina) at the Center for Cancer Research Sequencing Facility, National Cancer Institute. The libraries pooled from plates 1–4 were sequenced using one NovaSeq 6000 S4 Reagent Kit (200 cycles) and NovaSeq XP 4-Lane Kit (Illumina) with 100 bp paired-end reads as the sequencing parameter. The library pool from plate 5 was sequenced using the NovaSeq 6000 SP Reagent Kit (300 cycles; Illumina) with 150 bp paired-end reads as the sequencing parameter.

Moreover, after quality control, 11 samples were resequenced as plate 6 on a NextSeq 500 instrument using a NovaSeq 6000 S4 Reagent Kit (200 cycles) with 100 bp paired-end reads as the sequencing parameter. Technical replicates were placed on each plate to control for plate variability.

CITE-seq

Single-cell CITE-seq processing

Frozen PBMC samples were thawed, recovered and washed using RPMI medium with 10% FBS and 10 mg ml^{-1} DNase I (STEMCELL) and then processed as previously described[15] for CITE-seq staining. In brief, samples from different donors were pooled and different timepoints from the same donor were pooled separately so that each pool contains only one timepoint from one donor. PBMC pools were Fc blocked (Human TruStain FcX, BioLegend) and stained with Totalseq-C human 'hashtag' antibodies (BioLegend), washed with CITE-seq staining buffer (2% BSA in PBS). Hashtagged PBMC pools were then combined, and cells were stained with a cocktail of TotalSeq-C human lyophilized panel (BioLegend) of 137 surface proteins (including 7 isotype controls; Supplementary Table 11) and SARS-CoV-2 S1 protein probe. Cells were then washed, resuspended in PBS and counted before proceeding immediately to the single-cell partition step.

Single-cell CITE-seq library construction and sequencing

PBMC samples were partitioned into single-cell gel-bead in emulsion (GEM) mixed together with the reverse transcription mix using the 10x 5′ Chromium Single Cell Immune Profiling Next GEM v2 chemistry kit (10x Genomics), as previously described[15]. The reverse transcription step was conducted in the Veriti Thermal Cycler (Thermo Fisher Scientific). Single-cell gene expression, cell surface protein, T cell receptor (TCR) and B cell receptor (BCR) libraries were prepared according to the 10x Genomics user guides (https://www.10xgenomics.com/resources/user-guides/). All libraries were quality-controlled using the Bioanalyzer (Agilent) and quantified using Qubit fluorometric quantification (Thermo Fisher Scientific). 10x Genomics 5′ single-cell gene expression, cell surface protein tag, TCR and BCR libraries were pooled and sequenced on the Illumina NovaSeq platform (Illumina) using the following sequencing parameters: read1-100-cycle, i7-10-, i5-10, read2-100.

Serum isolation and protein characterization

Serum was collected directly in serum separator tubes and allowed to clot at room temperature for a minimum of 30 min. Within 2 h of blood collection, the tubes were centrifuged 1,800*g* for 10 min at room temperature. The top (serum) layer was removed using a pipette and stored in individual vials at −80 °C. Serum proteins were analysed using the Olink Target 96 Immuno-Oncology and Olink Target 96 Inflammation panels (Olink Proteomics, Uppsla, Sweden), which comprised 92 proteins each and uses the methodology based on the proximity extension assay. Data are reported as normalized protein expression (NPX) unit.

CBCs and lymphocyte phenotyping

For the participants, standard complete blood counts with differential (CBCs) were performed at the NIH CC in the Department of Laboratory Medicine. Lymphocyte (T cell, B cell, NK cell) flow cytometry quantification was performed using the BD FACSCanto II flow cytometer (BD Biosciences).

PBMC in vitro stimulation

PBMCs were thawed and cultured in PRMI1640 containing 10% fetal bovine serum, 2 mM glutamine, 0.055 mM beta-mercaptoethanol, 1% penicillin–streptomycin, 1 mM sodium pyruvate, 10 mM HEPES and 1% non-essential amino acids, and stimulated under the following conditions: (1) IL-15 (10 ng ml^{-1}), IL-12 (20 ng ml^{-1}), IL-18 (20 ng ml^{-1}) for 48 h; (2) IL-15 (50 ng ml^{-1}) for 48 h; (3) IL-18 (50 ng ml^{-1}) for 48 h; (4) IL-12 (20 ng ml^{-1}), IL-18 (20 ng ml^{-1}) for 48 h; (5) anti-CD3 (1 μg ml^{-1}), anti-CD28 (1 μg ml^{-1}) for 24 h; (6) non-stimulated controls. Protein Transport Inhibitor (BD Biosciences, 554724) and Brefeldin A (BFA, Invitrogen, 00-4506-51) were added 4 h before collection. The following cytokines were purchased from BioLegend: IL-15 (570304), IL-12 (573004) and IL-18 (592104).

Flow cytometry

B cell phenotyping panel including influenza HA probes

Thawed PBMCs were washed in RPMI culture medium containing 50 U ml^{-1} benzonase nuclease and then washed with PBS. Cells were incubated with LIVE/DEAD Fixable Blue Dye (Life Technologies), which was used to exclude dead cells from analysis. Cells were incubated with fluorochrome-conjugated HAs for influenza B (B/Washington/02/2019 and B/Phuket/3073/2013 combined on the same fluorochrome), and influenza A H1 (A/Hawaii/70/2019) and H3 (A/Hongkong/2671/2019) and fluorochrome-conjugated antibodies against IgM, IgA, CD21, CD85J, FCRL5, CD20, IgG, CD38, CD14, CD56, CD3, CD27, CD71, CD19 and IgD for 30 min at 4 °C in the dark. The dyes and detailed information of antibodies in the panel (Sarah Andrews, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH) are summarized in Supplementary Table 12. After incubation with antibodies for 30 min, cells were washed twice with FACS buffer (0.1% BSA/PBS (pH 7.4)) and fixed in 1% paraformaldehyde. A total of 5 million cells were acquired on the Cytek Aurora spectral cytometer (Cytek Biosciences; SpectroFlo (v.2.2.0)). Data were analysed using FlowJo (v.10; BD Biosciences).

General immune phenotyping panel

Thawed PBMCs were washed in RPMI culture medium containing 50 U ml^{-1} benzonase nuclease and then washed with PBS. Cells were incubated with LIVE/DEAD Fixable Blue Dye (Life Technologies), which was used to exclude dead cells from analysis. Cells were washed in FACS staining buffer (1× PBS, 0.5% fetal calf serum, 0.5% normal mouse serum and 0.02% NaN_3) and incubated with Human Fc block reagent (BD Bioscience, 564220) at room temperature for 5 min. Cells were stained at room temperature for 10 min in the dark with fluorochrome-conjugated antibodies against CCR7, CCR6, CXCR5, CXCR3 and TCRgd. Cells were then stained with fluorochrome-conjugated antibodies against CD45RA, CD16, CD11c, CD56, CD8, CD123, CD161, IgD, CD3, CD20, IgM, IgG, CD28, PD-1, CD141, CD57, CD45, CD25, CD4, CD24, CD95, CD27, CD1c, CD127, HLA-DR, CD38, ICOS, CD21, CD19 and CD14 at room temperature for 30 min in the dark. Cells were washed twice with FACS staining buffer (1× PBS, 0.5% fetal calf serum, 0.5% normal mouse serum and 0.02% NaN_3) and fixed in 1% paraformaldehyde. Supplementary Table 13 shows the clones and information of the antibodies used in the phenotyping panel. A total of 1 million PBMCs were acquired using the Cytek Aurora spectral cytometer (Cytek Biosciences; SpectroFlo (v.2.2.0)). The frequency of major populations was analysed using FlowJo (v.10; BD Biosciences) on the basis of previously described manual gating strategies[62,63,64].

In vitro stimulation T cell panel

In vitro simulated PBMCs were collected and washed in PBS. Cells were incubated with Zombie UV Fixable Viability Dye (BioLegend) in the dark (at room temperature) for 20 min. Cells were then washed and incubated with Human TruStain FcX (BioLegend) for 10 min and subsequently with anti-CCR7 antibodies for 10 min. A cocktail of fluorochrome-conjugated antibodies against CD8, CD4, HLA-DR, CD69, CD45RA, CD11c, CD5, CD3, TCRVa7.2, CD45RO, CD56, CD122, CD158e/k (KIR3DL1/DL2), KIR2D, NKG2A, CD14, CD29 and GPR56 was added and cells were stained for 30 min in the dark. Cells were washed and fixed using the Fixation/Permeablization kit (BD Biosciences). The intracellular proteins IFNγ, TNF and Ki-67 were stained after fixation. The samples were collected using the BD FACSymphony flow cytometer (BD Biosciences)

and analysed using FlowJo (v.10). A list of the antibodies used in the panel is provided in Supplementary Table 14.

Data processing and transformation

Bulk RNA-seq data processing

Sequencing reads from plate 5 were adaptor- and quality-trimmed to 100 bp using Trimmomatic (v.0.38.0)[65] to match the read length of the other plates (resulting reads with less than 100 bp were discarded). Reads were then aligned to the human genome hg38 using the STAR (v.2.6.0b) aligner. Duplicate reads from PCR amplification were removed based on unique molecular identifiers using UMI-tools (v.0.5.3). Gene expression quantification was performed using the featureCounts[66] function from Subread package (v.1.6.2). Samples with less than 5 million assigned reads were resequenced and replaced. Reads were normalized and log transformed using limma voom[67]. Low-expressed genes, defined as having fewer than five samples with >0.5 counts per million reads, were removed. Prevaccination (days −7 and 0) samples from the same healthy control participants were considered to be replicates and were used to estimate latent technical factors using the RUVs function of the RUVSeq[68] R package (v.1.18). Four latent variables were included to derive normalized gene expression values used for visualization and when specifically noted. Variable genes based on intraparticipant variability of prevaccination samples in the healthy control individuals and across technical replicates were filtered out, resulting in a total of 10,017 remaining genes for downstream analyses.

CITE-seq data processing

Single-cell sample demultiplexing and preprocessing

Single-cell sequencing data were demultiplexed, converted to FASTQ format, mapped to the human hg19 reference genome and counted using the CellRanger (10x Genomics) pipeline. The sample-level demultiplex was performed based on two levels as previously described[15]: (1) hashtag

antibody staining to distinguish different timepoint samples from a same participant; (2) single-nucleotide polymorphisms (SNPs) called from the whole-blood RNA-seq data to identify different participants. Specifically, CellRanger (v.6.0.1) was used for generating count matrix and the software package demuxlet (v.2, from the popscle software suite)[69] was used to match single-cell gene expression data to each donor and identify empty droplets and doublets.

Single-cell data clustering and cell annotation

Single-cell data were further processed using Seurat (v.4.0.3) running in R v.4.1.1. We removed cells with less than 200 and greater than 5,000 detected genes; greater than 60% of reads mapped to a single gene; greater than 15% mitochondrial reads; cell surface protein tag greater than 20,000; and hashtag antibody counts greater than 20,000. The protein data were normalized and denoised using the DSB method (v.0.3.0)[70]. The following parameters were used in the dsb normalization function: define.pseudocount = TRUE, pseudocount.use = 10, denoise_counts = TRUE, use.isotype.control = TRUE. The DSB-normalized protein data were used to generate the top variable features ($n = 100$) and principal components (PCs). The shared nearest neighbour graph followed by k-nearest neighbours clustering were then built using the FindNeighbors and FindClusters functions using the first 15 PCs in Seurat (v.4.0.3), respectively. Cell clusters were quality-controlled on the basis of their nearest neighbours and cell surface proteins. Cells were then further clustered within each major cell population using weighted-nearest neighbour (WNN) analysis in Seurat[71] (v.4.1.0) by integrating both cell surface protein and gene expression modalities. WNN FindMultiModalNeighbors was performed using both the top 10 PCs for cell surface protein and RNA of variable features. The WNN clusters were manually annotated and quality-controlled using the surface protein together with gene expression.

$CD8^+$ EM cell annotation for CITE-seq clusters

All $CD8^+$ cells were clustered using WNN as described above. $CD8^+$ clusters were annotated on the basis of their surface markers as reported[72]

together with gene expression profile. RNA expression of CD8$^+$ cells was mapped to an external dataset using the Seurat Label transfer method[71,73] (v.4.1.0). Clusters annotated as CD8$^+$ EM are surface CD45RO^{high}, CD45RA^{low}, CD95$^+$, CD62L^{low} and *CCR7* mRNA$^-$ with most cells (around 90%) mapped to CD8$^+$ EM phenotype cells in an external dataset[71,73].

Single-cell TCR data processing

CellRanger (v.6.0.1) was used to assemble V(D)J contigs. The V(D)J assignment and clonotype were from the CellRanger output of the filtered contig_annotations.csv file for each 10x lane. The data were combined for all lanes and paired TCRα and TCRβ chains for each single cell were combined using the scRepertoire R package (v.1.4.0)[74] and integrated with the single-cell CITE-seq Seurat object metadata. Cells annotated as CD8$^+$ T cells and with both α and β chains detected were filtered and analysed. CD8$^+$ subsets and GPR56$^+$CD8$^+$ EM cell clonality were visualized by Circos plots using the Circlize R package (v.0.4.14)[75]. For the purpose of visualization, cells from each subset were downsampled with equal numbers in each subset (a comparison between subsets is shown in Extended Data Fig. 4g) or in each timepoint (a comparison between timepoints is shown in Extended Data Fig. 4h,i). Cells were considered to be the same clone when they had identical CDR3 (both α and β chains). Identical clones were connected within each sample or each participant across timepoints with lines.

OLINK serum proteomics

Missing values were imputed using the *k*-nearest neighbours approach with $k = 10$ using the impute R package[76] (v.1.60.0). For each sample, probes targeting the same protein were averaged.

Cytek flow cytometry

Cell frequencies were generated by converting cell counts as fraction of live cells or lymphocytes as specified. The frequency data were $\log_2$-transformed for linear modelling. For populations with zero counts in any of the samples,

an offset equal to half of the smallest non-zero value was added across samples.

CBCs and lymphocyte phenotyping

Both absolute and relative counts were $\log_2$-transformed for linear modelling. Missing values were imputed using *k*-nearest neighbours approach. For parameters with zero values in any of the samples, an offset equal to half of the smallest non-zero value was added across samples.

Statistical analysis

Baseline differential expression analysis

Using the dream[77] function in the variancePartition R package (v.1.16.1), mixed-effects models were applied to determine differential levels of analytes (that is, whole-blood gene expression, serum proteins, cell frequencies, flu titre and SPR, and haematological parameters) between participants who had recovered from COVID-19 and healthy control participants in a sex-specific manner as follows: ~ 0 + group:sex + age + race + batch.effects + (1|participant.ID).

Batch-effect-related covariates were added to specific models depending on the assay type. For bulk RNA-seq, these include the four latent technical factors (see the 'Bulk RNA-seq data processing' section) and the timepoint-matched % neutrophils parameter from the CBC panel. For the Cytek and Olink platforms, sampling batch/plate was included as covariates. In addition to day 0, available samples from day −7 (in the RNA-seq and CBC panel) were included as baseline replicates in the modelling.

Sex-specific group differences were computed from the contrasts covid.Female − healthy.Female and covid.Male − healthy.Male. Overall COVID-19 versus healthy control difference was determined by combining the two contrasts, that is, (covid.Female − healthy.Female)/2 + (covid.Male − healthy.Male)/2. Sex difference linked to SARS-CoV-2 infection was derived from the contrast (covid.male − covid.female) − (healthy.male − healthy.female) to account for normal differences between males and

females. *P* values were adjusted for multiple testing within each assay type and contrast combination using the Benjamini–Hochberg method[78].

Association with TSD

To evaluate whether any of the differences detected at baseline had stabilized or might still be resolving, a linear model was used to test the association of relevant parameters with the time since COVID-19 diagnosis (TSD) among participants who had recovered from COVID-19: ~ 0 + sex + sex:scale(TSD) + age + race + (1|participant.ID).

Two asymptomatic participants without a known TSD were excluded from the model. Association was assessed separately for female and male individuals, and jointly by the combined contrast (female:TSD + male:TSD)/2. Dependent variables were converted to ranks in the model to reduce the effect of potential outliers.

Using a conservative approach, genes were classified as TSD-associated if they had an unadjusted $P < 0.05$ and were excluded from subsequent analyses as specified. To determine whether any of the baseline differential gene sets were associated with TSD, LEG modules were derived from the union of all LEGs of the same gene set from different contrasts (see the 'Bulk RNA-seq gene set module scores' section). A gene set was considered to be stable if none of three contrasts tested in the association model were significant (using an unadjusted *P* value threshold of 0.05).

Post-vaccination differential expression analysis

Similar to the workflow used in the baseline differential expression analysis, mixed-effects models were created to evaluate changes and group differences at each available timepoint after vaccination. Participants aged 65 and above were excluded as they received a different type of vaccine. In addition to the baseline covariates, the model also accounts for the participants' flu vaccination history within last 10 years as follows: ~ 0 + visit:group:sex + age + race + flu.vax.count.10yr + batch.effects + (1|participant.ID).

Three types of comparisons were examined using this model:

Timepoint-specific group differences: similar to the contrasts in the baseline model, but for individual timepoints post vaccination (day 1 to day 100).

Vaccine-induced changes in group difference: similar to the timepoint-specific contrasts above, but additionally subtracting off the corresponding baseline contrast to assess the changes relative to the baseline. For example, the differences in vaccine-induced changes for female individuals with COVID-19 versus healthy control individuals at D1 is evaluated with the contrast: (D1.covid.female − D1.healthy.female) − (baseline.covid.female − baseline.healthy.female).

Reversal of COVID-19 versus healthy control difference: instead of using the healthy control participants at the same corresponding timepoints as the reference, post-vaccination samples from the participants who had recovered from COVID-19 were compared to baseline healthy control individuals with the contrasts [timepoint].covid.female − baseline.healthy.female and [timepoint].covid.male − baseline.healthy.male. These contrasts can inform whether any prevaccination differences observed in the participants who had recovered from COVID-19 were reverted towards healthy baseline levels after vaccination. Reversal is defined as having a smaller absolute effect size (using the z.std value from the dream function) at D1 and D28 after vaccination compared with the baseline absolute effect size.

P values were adjusted for multiple testing for each timepoint, assay type and contrast combination using the Benjamini–Hochberg method.

Gene set enrichment of differentially expressed genes

Enriched gene sets were identified using the preranked GSEA algorithm implemented in the clusterProfiler R package (v3.17.0)[79]. Genes were ranked using signed $-\log_{10}$-transformed *P* values from differential expression models. Enrichment was assessed with gene set lists from MSigDB's Hallmark collection[80], blood transcriptomic modules[81] and cell type gene signatures[53]. Only gene sets with 10 to 300 genes were considered. *P* values were adjusted per gene set list for each contrast using

the Benjamini–Hochberg method and gene sets with FDR-adjusted $P < 0.05$ were considered to be significant. Baseline enriched gene sets were derived by intersecting significant gene sets extracted from differential expression models using samples independently from day –7, day 0, and both days combined. Genes associated with TSD at the baseline (see the 'Association with TSD' section; Supplementary Table 1) were excluded from the post-vaccination enrichment analyses to help to segregate the effect of vaccination from the natural temporal resolution of the SARS-CoV-2 infection.

Pseudobulk differential expression and GSEA

Single cells from a given sample were computationally pooled according to their cell type assignment by summing all reads for a given gene. Pseudobulk libraries that were made up by few cells and were therefore probably not modelled properly using bulk differential expression methods were removed from the analysis for each cell type to remove samples that contained fewer than 4 cells and with less than 35,000 library size after pooling. Low-expressed genes were removed for each cell type individually using the filterByExpr function of edgeR (v.3.26.8)[82] with min.count = 2. Log-transformed counts per million (CPM) of each gene were calculated with scaling factors for library size normalization provided by the calcNormFactors function. Differential expression analysis was performed using the same models described in the 'Post-vaccination differential expression analysis' section without running baseline models separately because the entire CITE-seq cohort was aged under 65 years. Batch assignment and the number of barcodes/cells per sample were included as batch effects in this model.

Similarly, GSEA was performed for each cell type in the same manner as described for the bulk RNA-seq data (see the 'Gene set enrichment of differentially expressed genes' section), which particularly focuses on the baseline enriched gene sets identified by the bulk RNA-seq analysis. The Monaco gene sets were excluded from the single-cell analysis given the cell clusters were annotated and no further cell type demultiplex needed.

Bulk RNA-seq gene set module scores

Gene set module scores were generated from RUVseq (v.1.18) normalized gene expression values (see the ‘Bulk RNA-seq data processing’ section) using the gene set variation analysis (GSVA) method in the GSVA R package (v.1.30.0)[83]. LEG module scores representing enriched pathway activities were calculated for relevant samples using LEGs identified by GSEA to enhance the signal-to-noise ratio. The average scores between days −7 and 0 were used for calculating post-vaccination changes relative to the baseline.

Pseudobulk gene set module score calculation

Module scores (gene set signature score) representing enriched pathway activities were calculated for each pseudobulk sample of certain cell types. The pseudobulk gene counts were corrected using the removeBatchEffect function in the limma package (v.3.42.2) to remove experimental batch and cell number effects and then normalized with voom[84]. The scores were then generated using the gene set variation analysis (GSVA) method from the GSVA R package (v.1.42.0)[83]. Specifically, for monocyte signatures, LEGs of BTM modules M4.0 and M11.0 were identified by GSEA from the (1) D0.COVR-F versus D0.HC-F and (2) D0.COVR-M versus D0.HC-M models. The union of LEGs was used for the score calculation for female and male samples.

For the BTM-M7.3 T cell activation signature and other signatures from acute COVID-19 data as indicated in the figures, LEGs were used from the indicated comparison groups for the score calculation of female and male individuals separately.

For the monocyte antigen presentation signature, the module score was generated using LEGs from the BTM-M71 enriched in antigen presentation (I) and M95.0 enriched in antigen presentation (II) gene sets of the comparison: D1 − D0 change between the COVR-M versus HC-M groups (Fig. 2f).

For the Hallmark IFNγ response module score, all genes from the gene set were used for calculating module scores in each cell type, so that the differences between cell types could be compared.

Single-cell module score calculation and visualization

To visualize the difference between subject groups in certain gene signatures using single data, the genes from the indicated gene sets were used to calculate the corresponding module score of each single cell. Module scores were calculated using the AddModuleScore function in Seurat (v.4.1.0) and then visualized in UMAP plots. For D1 versus D0 Hallmark IFNγ response module score difference (D1–D0) shown in UMAP projections (Fig. 2d), cells from the D1.HC-F, D1.COVR-F, D1.HC-M and D1.COVR-M groups were downsampled to the same number of cells. The UMAP embeddings of cells coloured with the average difference (D1–D0) of each high-resolution cell subsets are shown (each of the major cell clusters shown can contain one or more high-resolution cell subsets).

Single-cell module score calculation and test of external acute COVID-19 single-cell CITE-seq data

Single-cell data from the Brescia cohort of ref. [15] were downloaded from the Gene Expression Omnibus (GEO). Single monocyte data were extracted and single-cell data from the Brescia cohort were pooled as described in the 'Pseudobulk differential expression and GSEA' section. The gene set module scores of BTM modules M4.0 and M11.0 for all of the samples were generated using the union LEGs of male and female in the 'Gene set module score calculation' section. The pseudobulk gene counts were normalized using the varianceStabilizingTransformation function of the DEseq2 R package (v.1.34.0)[85]. The scores were then generated using the GSVA method from the GSVA R package (v.1.42.0)[83]. Given there are multiple samples from each participant, the differences between patient groups (healthy control, less severe and more severe, corresponding to HC, DSM-low and DSM-high in ref. [15]) were tested using the Limma (v.3.50.1) linear model, where samples from the same donors were treated as duplicates using duplicateCorrelation. *P* values of *t* statistics from the linear model of the indicated contrasts are shown.

Visualization of gene expression in heat maps

Heat maps showing pseudo-bulk data were generated using the ComplexHeatmap R package (v.2.10.0)[86]. The log[CPM]-normalized expression for each sample for a given cell type was calculated by pooling cells as described in the ‘Pseudobulk differential expression and GSEA’ section. Heat maps show the *z*-score of the normalized expression for each gene in each sample.

Data visualization

Plots were created using ggplot2 (v.3.3.5) with ggpubr (v.0.4.0) for statistical calculation unless noted.

End-point association

To evaluate the association of relevant parameters, including gene set module scores and cell frequencies, with IFN or antibody titre fold change end points, the following model was applied: end point ~ group:sex + scale(parameter):group:sex + age + race + flu.vax.count.10yr.

The end-point values were converted to rank to reduce the effects of potential outliers. Replicates from the same participants were averaged.

Serology

Influenza antibody titres below the detection limit of 1:20 were set to 1:10. Maximum titre across strains was calculated by normalizing titre levels across all of the samples from both D0 and D28 individually for each of the four strains followed by taking the maximum standardized titre for each sample.

Baseline titre difference analysis

For each of the four strains, a linear model was applied to determine the baseline titre differences between participants who had recovered from COVID-19 and healthy control participants in a sex-specific manner as follows: day 0 titre ~ group:sex + age + race.

Titre values were $\log_{10}$-transformed in the model, and sex-specific group differences were computed from the contrasts covid.Female − healthy.Female and covid.Male − healthy.Male. Participants aged 65 and above were excluded from the analysis.

D28 titre difference analysis

For post-vaccination titre response, influenza vaccination history and baseline titre were included as covariates to partly account for previous exposure, similar to the approach used for influenza vaccine evaluation by the Food and Drug Administration (page 27 of https://www.fda.gov/media/135687/download). Both D28 titre and D28/D0 FC were evaluated as end points to determine group differences between participants who had recovered from COVID-19 and healthy control participants for each of the four strains: endpoint ~ group:sex + age + race + flu.vax.count.10yr + day 0 titre.

For D28 FC, a negative binomial model with log link was applied using the glm.nb function in the MASS R package (v.7.3-53). A linear model was used to fit the D28 titres. Strain-specific titre values were $\log_{10}$-transformed in the model. Group differences were assessed using the same participants and contrasts as in the baseline analysis.

Influenza antibody avidity as measured using SPR was analysed in the same manner as the titre data across HA1 and HA2, with the exception that that a linear model was applied for the fold changes.

Concordance in the natural influenza infection cohorts

A prospective cohort study with participants profiled before and at least 21 days after natural influenza infection in two seasons[18] was used to assess

the residual effects of the infection separately in male and female individuals. Gene expression data were downloaded from the GEO (GSE68310). Participants with only influenza A virus infection ($n = 51$ female and $n = 35$ male) were identified and included for this analysis. Low-expressed probes were removed, and the remaining data were converted to gene-based expressions. No additional processing steps were performed as the data were already normalized.

Separately for each season, differential expression analysis between baseline (pre-infection) and spring (long-term post-infection) samples from the same individuals were performed using the dream function in the variancePartition R package (v.1.16.1). A mixed-effects model accounting for flu vaccination history and disease severity (based on fever grade: none, low and high) was constructed as follows: ~ 0 + timepoint:sex + age + num.flu.vaccination + fever.grade + (1|participant.ID).

Differentially expressed genes were identified using the contrasts Spring.F − Baseline.F and Spring.M − Baseline.M for female and male individuals, respectively. Sex difference was evaluated by the contrast (Spring.M − Baseline.M) − (Spring.F − Baseline.F). Concordance of differential expression results between the two seasons was evaluated on the basis of the correlation of effect size across genes (z.std values generated by dream).

Enrichment analysis was performed to determine whether the same set of genes was differentially expressed between pre- and post-influenza infection from this independent cohort and in participants who had recovered from COVID-19 compared with healthy control individuals before vaccination. To better match the age range of participants between the two studies, baseline differential gene analysis was performed again with participants under 65 years of age in the COVID-19 cohort (see the ‘Baseline differential expression analysis’ section). Given that the male participants showed stronger concordance between the two flu seasons (Extended Data Fig. 2b), COVID-19 differentially expressed genes were ranked by signed $-\log_{10}$-transformed P values and tested against a gene set formed by the intersect of differentially expressed ($P < 0.05$) genes in male participants from the flu infection cohort.

Elastic net multivariate predictive modelling

Elastic net models were constructed using the eNetXplorer R package (v.1.1.3)[87] to predict day 1 (D1) INFγ response after influenza vaccination with both CITE-seq and flow cytometry cell frequencies at D0 as predictors. A total of 33 participants (COVR-F = 11, HC-F = 8, COVR-M = 9, HC-M = 5) with both CITE-seq and flow cytometry data were included. On the basis of 20 runs of fivefold cross-validation, a grid of regularization parameters (α and λ) were tested to determine models with best performance and cell subsets with consistent predictive power. Model performance was assessed on the basis of the mean squared error between the predicted and observed response. The importance of a cell population was determined by the frequency that it was selected by the models (that is, having non-zero coefficient). *P* values of the model performance and feature importance were derived by comparing to null models constructed with permuted response.

TCR diversity metric calculation

Shannon's entropy (H' index) was calculated as a measure of TCR diversity[88,89]. Samples for each $CD8^+$ subset with fewer than 50 cells were filtered out from the calculation. All of the samples were downsampled to 50 cells because the diversity metric can be affected by the sample cell numbers. The process was repeated 1,000 times with random downsampling of 50 cells, and the median Shannon's index was used as an estimate of diversity for a given sample. Differences in the diversity metric between different $CD8^+$ subsets or timepoints were tested using two-tailed Wilcoxon tests.

Reversal genes and bootstrapping to infer the significance of difference in the reversal of monocyte-repressed signature between the COVR-F and COVR-M groups

Reversal genes are defined as those genes of which the COVID-19-recovered versus D0 healthy control absolute effect size (z.std values from

dream; see the ‘Post-vaccination differential expression analysis’ section) are smaller at both D1 and D28 compared with at D0.

Bootstrapping was used to determine the significance of the difference between the COVR-F and COVR-M groups in their proportion of baseline LEGs from the monocyte-depressed signature (BTM M4.0 and M11.0) that moved towards the healthy control baseline. Members from each participant group were randomly sampled with replacement in each round of the bootstrapping and their samples were analysed as described in the ‘Post-vaccination differential expression analysis’ section. The proportion of LEGs reversed after vaccination was calculated in each round for the COVR-F and COVR-M groups in classical and non-classical monocytes, separately, and the *P* values plotted in Fig. 4e were determined on the basis of 20 rounds of this procedure.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed data from the whole-blood bulk RNA-seq and single-cell CITE-seq are available from the NCBI GEO under accession numbers GSE194378 and GSE206265, respectively. Additional datasets, including clinical, proteomics, flow cytometry and influenza antibody measurements, are available at Zenodo (https://doi.org/10.5281/zenodo.5935845). The influenza infection dataset that we used was downloaded directly from the GEO (GSE68310).

Code availability

Analysis code, extended patient and sample metadata are available at GitHub (https://github.com/niaid/covid-flu).

References

1. Mina, M. J. et al. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* **366**, 599–606 (2019).

2. Tomalka, J. A., Suthar, M. S., Diamond, M. S. & Sekaly, R. P. Innate antiviral immunity: how prior exposures can guide future responses. *Trends Immunol.* **43**, 696–705 (2022).

3. Tsang, J. S. et al. Improving vaccine-induced immunity: can baseline predict outcome? *Trends Immunol.* **41**, 457–465 (2020).

4. Goodridge, H. S. et al. Harnessing the beneficial heterologous effects of vaccination. *Nat. Rev. Immunol.* **16**, 392–400 (2016).

5. Aaby, P., Netea, M. G. & Benn, C. S. Beneficial non-specific effects of live vaccines against COVID-19 and other unrelated infections. *Lancet Infect. Dis.* **23**, e34–e42 (2022).

6. Netea, M. G. et al. Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020).

7. Nalbandian, A. et al. Post-acute COVID-19 syndrome. *Nat. Med.* **27**, 601–615 (2021).

8. Olsen, S. J. et al. Changes in influenza and other respiratory virus activity during the COVID-19 pandemic—United States, 2020–2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 1013–1019 (2021).

9. Sudre, C. H. et al. Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).

10. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

11. Ursin, R. L. & Klein, S. L. Sex differences in respiratory viral pathogenesis and treatments. *Annu. Rev. Virol.* **8**, 393–414 (2021).

12. Wheatley, A. K. et al. Evolution of immune responses to SARS-CoV-2 in mild-moderate COVID-19. *Nat. Commun.* **12**, 1162 (2021).

13. Schultze, J. L. & Aschenbrenner, A. C. COVID-19 and the human innate immune system. *Cell* **184**, 1671–1692 (2021).

14. Sette, A. & Crotty, S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* **184**, 861–880 (2021).

15. Liu, C. et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell* **184**, 1836–1857 (2021).

16. Paludan, S. R. & Mogensen, T. H. Innate immunological pathways in COVID-19 pathogenesis. *Sci. Immunol.* **7**, eabm5505 (2022).

17. Reyes, M. et al. Plasma from patients with bacterial sepsis or severe COVID-19 induces suppressive myeloid cell production from hematopoietic progenitors in vitro. *Sci. Transl. Med.* **13**, eabe9599 (2021).

18. Zhai, Y. et al. Host transcriptional response to influenza and other acute respiratory viral infections—a prospective cohort study. *PLoS Pathog.* **11**, e1004869 (2015).

19. Pulendran, B. Systems vaccinology: probing humanity's diverse immune systems with vaccines. *Proc. Natl Acad. Sci. USA* **111**, 12300–12306 (2014).

20. Tsang, J. S. Utilizing population variation, vaccination, and systems biology to study human immunology. *Trends Immunol.* **36**, 479–493 (2015).

21. Pulendran, B., S. Arunachalam, P. & O'Hagan, D. T. Emerging concepts in the science of vaccine adjuvants. *Nat. Rev. Drug Discov.* **20**, 454–475 (2021).

22. Bucasas, K. L. et al. Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *J. Infect. Dis.* **203**, 921–929 (2011).

23. Auladell, M. et al. Influenza virus infection history shapes antibody responses to influenza vaccination. *Nat. Med.* **28**, 363–372 (2022).

24. Truong, K.-L. et al. Killer-like receptors and GPR56 progressive expression defines cytokine production of human $CD4^+$ memory T cells. *Nat. Commun.* **10**, 2263 (2019).

25. Nicolet, B. P. et al. CD29 identifies IFN-γ-producing human $CD8^+$ T cells with an increased cytotoxic potential. *Proc. Natl Acad. Sci. USA* **117**, 6686–6696 (2020).

26. Bergamaschi, L. et al. Longitudinal analysis reveals that delayed bystander $CD8^+$ T cell activation and early immune pathology distinguish severe COVID-19 from mild disease. *Immunity* **54**, 1257–1275 (2021).

27. Bangs, S. C. et al. Human $CD4^+$ memory T cells are preferential targets for bystander activation and apoptosis. *J. Immunol.* **182**, 1962–1971 (2009).

28. White, J. T., Cross, E. W. & Kedl, R. M. Antigen-inexperienced memory $CD8^+$ T cells: where they come from and why we need them. *Nat. Rev. Immunol.* **17**, 391–400 (2017).

29. Maurice, N. J., Taber, A. K. & Prlic, M. The ugly duckling turned to swan: a change in perception of bystander-activated memory CD8 T cells. *J. Immunol.* **206**, 455–462 2021).

30. Abers, M. S. et al. An immune-based biomarker signature is associated with mortality in COVID-19 patients. *JCI Insight* **6**, 144455 (2021).

31. Herndler-Brandstetter, D. et al. Post-thymic regulation of CD5 levels in human memory T cells is inversely associated with the strength of responsiveness to interleukin-15. *Hum. Immunol.* **72**, 627–631 (2011).

32. Jacomet, F. et al. Evidence for eomesodermin-expressing innate-like $CD8^+$ KIR/$NKG2A^+$ T cells in human adults and cord blood samples. *Eur. J. Immunol.* **45**, 1926–1933 (2015).

33. Jin, J.-H. et al. Virtual memory CD8$^+$ T cells restrain the viral reservoir in HIV-1-infected patients with antiretroviral therapy through derepressing KIR-mediated inhibition. *Cell. Mol. Immunol.* **17**, 1257–1265 (2020).

34. Yang, J., Zhu, H., Murphy, T. L., Ouyang, W. & Murphy, K. M. IL-18-stimulated GADD45β required in cytokine-induced, but not TCR-induced, IFN-γ production. *Nat. Immunol.* **2**, 157–164 (2001).

35. Kastenmüller, W., Torabi-Parizi, P., Subramanian, N., Lämmermann, T. & Germain, R. N. A spatially-organized multicellular innate immune response in lymph nodes limits systemic pathogen spread. *Cell* **150**, 1235–1248 (2012).

36. Wimmers, F. et al. The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. *Cell* **184**, 3915–3935 (2021).

37. Debisarun, P. A. et al. Induction of trained immunity by influenza vaccination—impact on COVID-19. *PLoS Pathog.* **17**, e1009928 (2021).

38. Arunachalam, P. S. et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).

39. Schulte-Schrepping, J. et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440 (2020).

40. Pérez-Gómez, A. et al. Dendritic cell deficiencies persist seven months after SARS-CoV-2 infection. *Cell. Mol. Immunol.* **18**, 2128–2139 (2021).

41. Phetsouphanh, C. et al. Immunological dysfunction persists for 8 months following initial mild-to-moderate SARS-CoV-2 infection. *Nat. Immunol.* **23**, 210–216 (2022). 1–7.

42. Utrero-Rico, A. et al. Alterations in circulating monocytes predict COVID-19 severity and include chromatin modifications still detectable six months after recovery. *Biomedicines* **9**, 1253 (2021).

43. You, M. et al. Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. *Nat. Cell Biol.* **23**, 620–630 (2021).

44. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).

45. Patel, A. A. et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017).

46. Cheong, J.-G. et al. Epigenetic memory of COVID-19 in innate immune cells and their progenitors. Preprint at *bioRxiv* https://doi.org/10.1101/2022.02.09.479588 (2022).

47. Tough, D. F., Borrow, P. & Sprent, J. Induction of bystander T cell proliferation by viruses and type I interferon in vivo. *Science* **272**, 1947–1950 (1996).

48. Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2020).

49. Scully, E. P. et al. Sex and gender differences in testing, hospital admission, clinical presentation, and drivers of severe outcomes from COVID-19. *Open Forum Infect. Dis.* **8**, ofab448 (2021).

50. Tsang, J. S. et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* **157**, 499–513 (2014).

51. Kotliarov, Y. et al. Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. *Nat. Med.* **26**, 618–629 (2020).

52. Nakaya, H. I. et al. Systems biology of vaccination for seasonal influenza in humans. *Nat. Immunol.* **12**, 786–795 (2011).

53. Monaco, G. et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640 (2019).

54. Harris, P. A. et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).

55. Harris, P. A. et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).

56. Ravichandran, S. et al. Antibody signature induced by SARS-CoV-2 spike protein immunogens in rabbits. *Sci. Transl. Med.* **12**, eabc3539 (2020).

57. Ravichandran, S. et al. Longitudinal antibody repertoire in 'mild' versus 'severe' COVID-19 patients reveals immune markers associated with disease severity and resolution. *Sci. Adv.* **7**, eabf2467 (2021).

58. Tang, J. et al. Antibody affinity maturation and plasma IgA associate with clinical outcome in hospitalized COVID-19 patients. *Nat. Commun.* **12**, 1221 (2021).

59. Khurana, S. et al. MF59 adjuvant enhances diversity and affinity of antibody-mediated immune response to pandemic influenza vaccines. *Sci. Transl. Med.* **3**, 85ra48 (2011).

60. Khurana, S. et al. Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. *Nat. Med.* **22**, 1439–1447 (2016).

61. Khurana, S. et al. Repeat vaccination reduces antibody affinity maturation across different influenza vaccine platforms in humans. *Nat. Commun.* **10**, 3338 (2019).

62. De Biasi, S. et al. Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia. *Nat. Commun.* **11**, 3434 (2020).

63. Heit, A. et al. Vaccination establishes clonal relatives of germinal center T cells in the blood of humans. *J. Exp. Med.* **214**, 2139–2152 (2017).

64. Park, L. M., Lannigan, J. & Jaimes, M. C. OMIP-069: forty-color full spectrum flow cytometry panel for deep immunophenotyping of major cell subsets in human peripheral blood. *Cytometry A* **97**, 1044–1051 (2020).

65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

67. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

68. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).

69. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).

70. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nat.*

Commun. **13**, 2099 (2022).

71. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).

72. Martin, M. D. & Badovinac, V. P. Defining memory CD8 T cell. *Front. Immunol.* **9**, 2692 (2018).

73. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

74. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. *F1000Res.* **9**, 47 (2020).

75. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

76. Hastie, T., Tibshirani, R., Narasimhan, B. & Chu, G. impute: imputation for microarray data. R package version 1.72.1 (2022).

77. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).

78. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

79. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).

80. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

81. Li, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat. Immunol.* **15**, 195–204 (2014).

82. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

83. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

84. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

86. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

87. Candia, J. & Tsang, J. S. eNetXplorer: an R package for the quantitative exploration of elastic net families for generalized linear models. *BMC Bioinformatics* **20**, 189 (2019).

88. Rosati, E. et al. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* **17**, 61 (2017).

89. Morris, E. K. et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* **4**, 3514–3524 (2014).

Acknowledgements

We thank the study participants, who helped us to expand our scientific understanding of COVID-19 and viral infections in general; K. Sadtler for assistance with the development of the Cytek 36-colour flow cytometry panel; Q. Xu for assistance with the design of the cytokine stimulation

experiments; F. Cheung for computational assistance; M. Mulè for providing software (dsb); the staff at the NCI CCR Sequencing Facility for sequencing support; the NIAID HPC Team for computing support; and R. Germain for reading the manuscript. Figs. 1a,b, 2a, 3a and 4a and Extended Data Fig. 2a were created using BioRender. This research was supported by the Intramural Research Programs of the NIAID and the Intramural Programs of the NIH Institutes supporting the NIH Center for Human Immunology. J.H. was funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract no. 75N91019D00024. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

Author information

Author notes

1. LaQuita Snow

   Present address: Johns Hopkins University, Baltimore, MD, USA

2. John S. Tsang

   Present address: Yale Center for Systems and Engineering Immunology and Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA

3. John S. Tsang

   Present address: Department of Biomedical Engineering, Yale University, New Haven, CT, USA

4. These authors contributed equally: Rachel Sparks, William W. Lau, Can Liu

Authors and Affiliations

1. Multiscale Systems Biology Section, Laboratory of Immune System Biology, NIAID, NIH, Bethesda, MD, USA

Rachel Sparks, William W. Lau, Can Liu, Kiera L. Vrindten, Guangping Sun, Milann Cox, Neha Bansal, Laura E. Failla, Andrew J. Martins & John S. Tsang

2. Graduate Program in Biological Sciences, University of Maryland, College Park, MD, USA

Can Liu

3. NIH Center for Human Immunology, NIAID, NIH, Bethesda, MD, USA

Kyu Lee Han, Galina Koroleva, Jinguo Chen, Amrita Mukherjee, Brian A. Sellers, Richard Apps & John S. Tsang

4. Division of Intramural Research, NIAID, NIH, Bethesda, MD, USA

Guangping Sun

5. Vaccine Research Center, NIAID, NIH, Bethesda, MD, USA

Sarah F. Andrews & Adrian B. McDermott

6. Division of Viral Products, Center for Biologics Evaluation and Research (CBER), FDA, Silver Spring, MD, USA

Jody Manischewitz, Gabrielle Grubbs, Lisa R. King, Juanjie Tang, Hana Golding & Surender Khurana

7. Laboratory of Clinical Immunology and Microbiology, NIAID, NIH, Bethesda, MD, USA

Stephanie Leimenstoll & LaQuita Snow

8. Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Evan M. Bloch

9. Clinical Center Nursing Department, NIH, Bethesda, MD, USA

Princess Barber, Daly Cantave, Anne Carmona, Alaina K. Magnani, Valerie Mohammed, Cindy Palmer & Deitra Shipman

10. Clinical Research Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

Jean Hammer

Consortia

OP11 Clinical Staff

- Princess Barber
- , Daly Cantave
- , Anne Carmona
- , Jean Hammer
- , Alaina K. Magnani
- , Valerie Mohammed
- , Cindy Palmer
- & Deitra Shipman

Contributions

Conceptualization: R.S. and J.S.T. Methodology: W.W.L., C.L. and J.S.T. Software and formal analysis: W.W.L. and C.L. Investigation: R.S., W.W.L., C.L., K.L.H., N.B., J.M., G.G., L.R.K., G.K., J.C., J.T. and A.J.M. Resources: R.S., W.W.L., C.L., K.L.V., G.S., M.C., S.F.A., N.B., L.E.F., S.L., L.S., P.B., D.C., A.C., J.H., A.K.M., V.M., C.P., D.S., A.M., B.A.S., R.A., A.B.M., E.M.B., H.G., S.K. and J.S.T. Data curation: R.S., W.W.L., C.L., K.L.V., M.C., N.B. and L.E.F. Writing—original draft: R.S. and J.S.T. Writing—review and editing: R.S., W.W.L., C.L. and J.S.T. Visualization: R.S., W.W.L. and C.L. Supervision: R.S., H.G., S.K. and J.S.T. Project administration: R.S. and J.S.T. Funding acquisition: R.S. and J.S.T. Human

study and sample acquisition: R.S., K.L.V., G.S., M.C., N.B., L.E.F., S.L., L.S., P.B., D.C., A.C., J.H., A.K.M., V.M., C.P., D.S. and A.M. Flow cytometry analysis of PBMCs (including influenza-specific cell analyses): K.L.H. with panel designs, reagents and help from S.F.A. Whole-blood RNA-seq analysis: G.K. and N.B. Influenza microneutralization titres, influenza antibody SPR and SARS-CoV-2 neutralization titres: J.M., G.G., L.R.K. and J.T. CITE-seq data generation: C.L., A.J.M. and J.C. with help from G.K. and K.L.V. PBMC cytokine stimulation experiments: C.L. Integrative analysis of flow cytometry, CITE-seq, blood RNA-seq, circulating protein and antibody data: W.W.L. and C.L. with conceptual contributions from J.S.T.

Corresponding author

Correspondence to John S. Tsang.

Ethics declarations

Competing interests

E.M.B. reports personal fees and non-financial support from Terumo BCT, Abbott Laboratories, Tegus and UptoDate, outside of the submitted work; is a member of the US FDA Blood Products Advisory Committee. Any views or opinions that are expressed in this manuscript are those of the author's, based on his own scientific expertise and professional judgement; they do not necessarily represent the views of either the Blood Products Advisory Committee or the formal position of FDA, and do not bind or otherwise obligate or commit either Advisory Committee or the Agency to the views expressed. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Andreas Wack and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are

available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Baseline differences between COVID-19-recovered participants and matching controls.

a, Box plot showing the distribution of time since diagnosis (TSD; x-axis) in COVR-F (n = 16) and COVR-M (n = 15). Two participants with asymptomatic COVID-19 infection and thus unknown TSD are not included. Significance of group difference is determined by two-tailed Wilcoxon test. **b**, Scatterplot showing the correlation between the TSD (x-axis) and the SARS-CoV-2 neutralization titre (WA1 strain; y-axis) for COVR individuals at D0 prior to influenza vaccination. Spearman's rank correlation and unadjusted *P* values are shown. Participants with asymptomatic COVID-19 infection not included in TSD analyses. The shaded area represents 95% confidence interval. **c**, Similar to (**b**), but for the percentage of monocytes in peripheral blood as measured by the complete blood count (y-axis) at D0. Dotted lines represent median level in HC-F and HC-M. **d**, Similar to (**b**), but for the proportion of CD11c+ dendritic cells (DCs; as fraction of live cells; y-axis) as measured by flow cytometry of PBMCs at D0. **e**, Blood transcriptomic analysis of the stable baseline (prevaccination) differences among COVR and HC groups. Enrichment plot shows the normalized enrichment scores (NES) of selected gene sets of the different comparisons (GSEA FDR < 0.05; Methods; see Supplementary Table 3 for all significant gene sets with FDR < 0.05). The NES are plotted separately for COVR-F versus HC females (HC-F), COVR-M versus HC males (HC-M), or the difference between the two sets of comparisons (COVR-M versus COVR-F taking healthy sex differences into account). Positive (negative) NES indicates that gene set scores are higher (lower) in the first group than the second group listed in the

comparison. Only gene sets not correlated with TSD across COVR individuals at baseline are considered stable. **f**, Similar to Fig. 1d, but for percent of CD3+ cells (T cells). **g**, Similar to (**e**), but for a subset of monocyte and T cell activation gene sets with significant enrichment ($P < 0.05$) using the D0 CITE-seq pseudobulk expression for the specified cell types (Methods; see Supplementary Table 5 for complete results). **h**, Scatterplots showing the relationship between the TSD and leading-edge gene (LEG) module scores [left two boxes: the T cell activation gene set (BTM-M7.3); right two boxes: the union of the LEGs from gene sets BTM M4.0 and M11.0; Methods] in COVR-F (n = 12) (top row) and COVR-M (n = 12) (bottom row) at D0 using the CITE-seq pseudobulk data of the indicated cell types. Each dot represents a COVR individual. The dotted lines represent the median score for the sex-matched HC group at D0 in the comparison shown. Spearman's rank correlation and *P* values are shown. The shaded area represents 95% confidence interval. **i**, (left) Box plot comparing the classical monocyte pseudobulk module scores of the LEGs used in Fig. 1f (union of female (F) and male (M) gene sets) in an acute COVID-19 CITE-seq dataset from Liu *et al*.[15]. Both M (n = 50) and F (n = 9) individuals are included in all three groups (HC n = 13, less severe n = 21, more severe n = 25). Each dot represents a sample. Unadjusted *P* values from the indicated two-group comparisons are shown. *P* values were generated using the moderated *t* statistics from a linear model in which samples from the same donors were treated as duplicates (Methods). (right) Bubble plot showing expression of the genes in Fig. 1f right panel within the classical monocyte CITE-seq data from Liu *et al*. in the same three patient groups shown in the left panel. All plot elements are the same as indicated in Fig. 1.

Extended Data Fig. 2 Persistent post-infection gene expression changes following natural influenza infection.

a, Schematic showing the approach used to evaluate changes in blood gene expression between before (pre-infection baseline) and months after natural influenza infection over two distinct seasons published in Zhai *et al*.[18], and how those gene changes may relate to sex-specific differences resulted from prior COVID-19 in this study. **b**, Scatter density plot showing the correlation between the gene expression changes (Supplementary Table 6)

before (fall) and after (spring) natural influenza A infection in 2009 (x-axis) and 2010 (y-axis) for females (F; left), males (M; centre), and M vs F contrast (right). Shown are Spearman's rank correlation and unadjusted *P* values. **c**, Gene set enrichment plot of the genes that increased in M between fall (pre-infection) and spring (post-infection) in both 2009–2010 and 2010–2011 seasons. Genes were ranked by the signed -$\log_{10}$-transformed (unadjusted) P values in the COVID-19-recovered (COVR)-M vs COVR-F contrast at baseline using only participants under 65 years of age. The tick marks denote the location of the genes in the influenza gene set. The diagram in **a** was created using BioRender.

Extended Data Fig. 3 Sex-specific molecular, cellular, and humoral response differences to influenza vaccination in COVID-19-recovered individuals and matching controls.

a, Similar to Extended Data Fig. 1e but here showing enriched gene sets in whole blood comparing the early (D1 and D7) influenza vaccination responses in COVID recovered (COVR) vs. healthy control (HC) participants for females (F; Contrast 1), males (M; Contrast 2), and sex differences (Contrast 2 vs. 1; i.e., COVR-M versus COVR-F taking healthy sex differences into account; see Methods). Plotted are the gene sets that show significant changes from the baseline (D-7 and D0) within each comparison group [e.g., COVR-F and HC-F for D1] and significant differences between the two groups at the indicated timepoints (FDR < 0.05; see Supplementary Table 5). **b**, Similar to Fig. 2e, but showing the D0 Hallmark IFNγ Response module score for the indicated cell types from the CITE-seq pseudobulk expression data. CD4 = $CD4^+$ T cells; cDC = conventional/myeloid dendritic cells; B = B cells. **c**, Box plot showing the D7 whole blood signature score from genes identified in Nakaya *et al.*[52] whose D7/D0 fold-change positively correlated with D28 influenza hemagglutination inhibition titres. Only participants under 65 years of age [COVR-F (n = 15), COVR-M (n = 14), and HC-F (n = 16), and HC-M (n = 14)] are included. Significance of differences is determined by two-tailed Wilcoxon rank-sum test. **d**, Scatter plot showing the correlation of the whole blood D1 – D0 Hallmark Interferon Gamma Response gene set module score (x-axis) to the whole blood D7 – D0 plasmablast (PB) gene

set module score (left y-axis; Monaco *et al*.[53]) and D7 – D0 difference of influenza-specific PB (all $HA^{+}CD27^{+}CD38^{+}CD20^{low}CD21^{low}$) frequency from flow cytometry (right y-axis; as fraction of $CD19^{+}$ B cells). Only study participants < 65 years of age are included. Spearman's rank correlation and unadjusted *P* values are shown. **e**, Box plots showing the D0 (prevaccination) microneutralization titres for each of the four strains in the seasonal influenza vaccine (columns) in females (COVR-F and HC-F) under the age of 65. Unadjusted *P* values are from linear models accounting for age and race (Methods). **f**, Similar to (**e**) but for males (COVR-M and HC-M) under 65 years of age. **g**, Maximum standardized influenza vaccine titre (among the four strains in the vaccine) at D28 after vaccination for females (left) and males (right), respectively. Unadjusted *P* values are from linear regression models accounting for age, race, influenza vaccination history, and baseline influenza titre (Methods). Unadjusted *P* values are shown. All plot elements are the same as indicated in Fig. 2.

Extended Data Fig. 4 GPR56^{+} virtual memory-like CD8^{+} T cells contribute to increased day 1 IFNγ response in COVID-19-recovered males.

a, Scatterplots showing the correlation between the day 0 (D0) $\log_2$ frequency of early effector-like $CD8^{+}$ T cells measured by flow cytometry (as fractions of live lymphocytes; see Population 50 in Supplementary Table 9 and Supplementary Fig. 1; x-axis) and the change (D1 – D0) in serum interferon gamma (IFNγ) protein levels measured by the OLINK platform (y-axis) for COVID-19-recovered (COVR) females (COVR-F; top, n = 14) and COVR males (COVR-M; bottom; n = 11). Spearman's rank correlation and *P* values are shown. **b**, Similar to **a** but showing the correlation between the D0 $\log_2$ frequency of early effector-like $CD8^{+}$ T cells measured by from flow cytometry (as fraction of live lymphocytes; x-axis) and the change (D1 - D0) in the whole blood signature score of the Hallmark IFNγ Response gene set (y-axis). **c**, Box plots comparing D0 and D1 pseudobulk IFNγ gene (*IFNG*) expression (y-axis) in $GPR56^{+}$ $CD8^{+}$ EM population for HC-F (n = 8), COVR-F (n = 12), HC-M (n = 8) and COVR-M (n = 12). Significance is determined by a linear model accounting for age, race, and

influenza vaccination history (Methods). **d**, Bar plot showing the *t* statistic of the vaccine-induced change (D1 - D0) in *IFNG* expression using CITE-seq pseudobulk data (x-axis) within the GPR56$^+$ and GRP56$^-$ CD8$^+$ EM for HC-F (n = 8), COVR-F (n = 12), HC-M (n = 8), and COVR-M (n = 12). * $P < 0.05$ with exact value shown in **(c)**. **e**, Scatter plot showing the correlation between GPR56$^+$ CD8$^+$ EM cell frequency (as fractions of total CD8$^+$ EM in the CITE-seq data; x-axis) and BTM-M7.3 T cell activation signature score of CD8$^+$ EM cells computed using CITE-seq pseudobulk gene expression data (y-axis). Spearman correlation and *P* values are shown. The shaded area represents the 95% confidence interval. **f**, Related to Fig. 3h but showing CD45RA and CD45RO only with CD8$^+$ TEMRA cells included as an additional comparator. **g**, (left) Circos plot of T cell receptor (TCR) clonality for different CD8$^+$ T cell subsets at D0. Segments in the outer circle represent different CD8$^+$ T cell populations. Segments in inner circle represent male (M) and female (F) for both COVR and HC participants. Grey lines connect clones sharing identical CDR3 sequences within each individual. Cell subsets are downsampled for visualization (Methods). (right) Box plot showing Shannon's entropy index (y-axis) at D0 for each of the indicated CD8$^+$ populations. Significance of differences is determined by two-tailed Wilcoxon test. Shannon's entropy index evaluates the TCR repertoire diversity for each sample. Higher indices indicate higher diversity (i.e., fewer shared clones shown in Circos plot). EM, effector memory; CM, central memory; TEMRA, EM cells re-expressing CD45RA. **h**, (left) Circos plot of TCR clonality for GPR56$^+$ CD8$^+$ EM cells at different timepoints. Segments in the outer circle represent different days in the study (D0, D1, D28). Segments in the inner circle represent males (M) and females (F) for both COVR and HC participants. Grey lines connect clones sharing identical CDR3 sequences within each sample. Timepoints are downsampled for visualization purposes (Methods). (right) Box plot showing Shannon's entropy index (y-axis) of TCR clonality at each of the indicated time points (D0, D1, D28; x-axis) for GPR56$^+$ CD8$^+$ EM T cells (left) and GPR56$^+$ CD8$^+$ TEMRA (right). Significance of differences is determined by two-tailed Wilcoxon test. **i**, (left) Similar to (**h**), but showing the shared clones among different timepoints (segments in the outer circle). Segments in the inner circle represent unique clones for each individual. Clones and lines connecting shared clones are coloured. Samples with less

than 30 cells were filtered out for visualization purposes. (right) Line chart showing frequencies of each clone (y-axis) shown in Circos plot (left) at D0, D1 and D28 for each participant. *P*-values of paired Wilcoxon test are shown comparing the clone frequency differences among D0, D1 and D28. **j**, Related to Fig. 3i but showing the frequencies of IFNγ$^+$ NK, IFNγ$^+$CD45RA$^+$CD45RO$^+$ TEMRA CD8$^+$ T cells and IFNγ$^+$ MAIT cells after IL-15 stimulation in vitro. All plot elements are the same as indicated in Fig. 3. Unadjusted *P* values are shown.

Extended Data Fig. 5 Changes in immune states in COVID-19-recovered individuals following influenza vaccination.

a, Distributions of gene-level difference of the innate immune receptor (IIR) signature (see Fig. 1f) in classical monocytes separately for females (F) and males (M) [shown as z-scores, on a per gene level, capturing the average difference between COVID Recovered (COVR) at the indicated timepoint (top to bottom: D0, D1, and D28) and healthy control (HC) at D0; see Methods]. Dashed red vertical lines represent the median of the distribution. Dark tick marks at the bottom represent individual genes and coloured dots highlight specific genes of interest. Significance of differences from D0 is determined by paired two-tailed Wilcoxon rank-sum test. **b**, Similar to (**a**) but for the non-classical monocytes (see Fig. 1g). **c**, Similar to Fig. 4d but for non-classical monocytes (see Fig. 1g for the innate receptor signature in non-classical monocytes). **d**, Similar to Fig. 4b but for COVR-F (red) and COVR-M (blue) only and gene sets shown on top of each plot. Box plots showing the classical monocyte LEG module scores (y-axis) of gene sets from Supplementary Fig. 2: antigen presentation related gene sets, Hallmark Inflammatory response, Hallmark TNF-α signalling via NF-κB, and MS-1 signature from Reyes *et al.*[17]. LEGs from the first three gene sets were found to be repressed in acute COVID-19 patients in Liu *et al.*[15]. **e**, Similar to (**d**), but for non-classical monocytes. All plot elements are the same as indicated in Fig. 4. Unadjusted *P* values are shown.

Extended Data Table 1 Cohort characteristics

Supplementary information

Supplementary Information

Supplementary Discussion, Supplementary References and Supplementary Figs. 1–7.

Reporting Summary

Peer Review File

Supplementary Table 1

Baseline parameter TSD association results (all participants).

Supplementary Table 2

Baseline parameter DE results (all participants).

Supplementary Table 3

Baseline whole blood GSEA results (all participants).

Supplementary Table 4

Significant ($P < 0.05$) single-cell timepoint DE results.

Supplementary Table 5

Timepoint whole-blood and single-cell GSEA results.

Supplementary Table 6

Zhai et al. natural influenza infection cohort DE results.

Supplementary Table 7

Post-influenza vaccination timepoint DE results.

Supplementary Table 8

Influenza vaccine antibody titre and SPR D0 and D28 model results.

Supplementary Table 9

Gated populations in the 36-colour Cytek panel.

Supplementary Table 10

COVR-M $CD8^+$ EM markers from CITE-seq.

Supplementary Table 11

The protein surface marker panel used for CITE-seq, BioLegend Total-seq-C Human Universal Cocktail, V1.0.

Supplementary Table 12

Antibodies and influenza probes used in the B cell flow panel.

Supplementary Table 13

Antibodies and viability dyes used in the 36-colour Cytek flow cytometry panel.

Supplementary Table 14

Antibodies and viability dyes used in T cell stimulation assays.

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

About this article

Cite this article

Sparks, R., Lau, W.W., Liu, C. *et al.* Influenza vaccination reveals sex dimorphic imprints of prior mild COVID-19. *Nature* **614**, 752–761 (2023). https://doi.org/10.1038/s41586-022-05670-5

- Received: 03 February 2022
- Accepted: 19 December 2022
- Published: 04 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05670-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

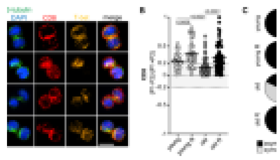
This article is cited by

- **Post-vaccination serum cytokines levels correlate with breakthrough influenza infections**

    - Weichun Tang
    - Hang Xie
    - Ewan P. Plant

  Scientific Reports (2023)

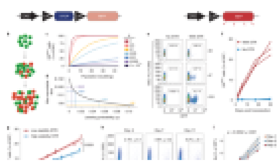
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05670-5



Asymmetric cell division shapes naive and virtual memory T-cell immunity during ageing

11 May 2021

Mariana Borsa, Niculò Barandun, … Annette Oxenius



Replicative history marks transcriptional and functional disparity in the CD8+ T cell memory pool

07 April 2022

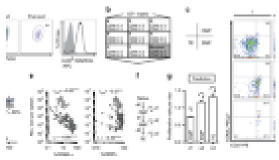
Kaspar Bresser, Lianne Kok, … Ton N. Schumacher



Building a T cell compartment: how immune cell development shapes function

03 June 2020

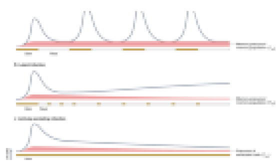
Miles P. Davenport, Norah L. Smith & Brian D. Rudd



Differential expansion of T central memory precursor and effector subsets is regulated by division speed

08 January 2020

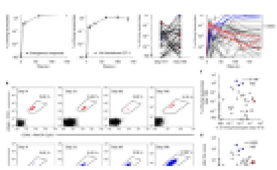
Lorenz Kretschmer, Michael Flossdorf, … Veit R. Buchholz



‘Stem-like’ precursors are the fount to sustain persistent CD8+ T cell responses

27 May 2022

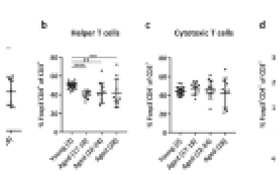
Dietmar Zehn, Robert Thimme, … Annette Oxenius



Early emergence of T central memory precursors programs clonal dominance during chronic viral infection

26 October 2020

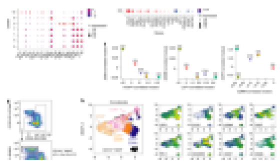
Simon Grassmann, Lorenz Mihatsch, … Veit R. Buchholz



[Response kinetics reveal novel features of ageing in murine T cells](#)

03 April 2019

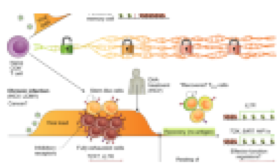
Daan K. J. Pieren, Noortje A. M. Smits, … Teun Guichelaar



[Two subsets of stem-like CD8+ memory T cell progenitors with distinct fate commitments in humans](#)

12 October 2020

Giovanni Galletti, Gabriele De Simone, … Enrico Lugli



[T cell exhaustion—a memory locked behind scars](#)

26 July 2021

Amir Yousif & Hazem E. Ghoneim

- Article
- Published: 18 January 2023

Functional T cells are capable of supernumerary cell division and

longevity

- Andrew G. Soerens,
- Marco Künzli,
- Clare F. Quarnstrom,
- Milcah C. Scott,
- Lee Swanson,
- JJ. Locquiao,
- Hazem E. Ghoneim,
- Dietmar Zehn,
- Benjamin Youngblood,
- Vaiva Vezys &
- …
- David Masopust

Show authors

Nature **volume 614**, pages 762–766 (2023)

- 74k Accesses
- 2 Citations
- 961 Altmetric
- Metrics details

Abstract

Differentiated somatic mammalian cells putatively exhibit species-specific division limits that impede cancer but may constrain lifespans[1,2,3]. To provide immunity, transiently stimulated $CD8^+$ T cells undergo unusually rapid bursts of numerous cell divisions, and then form quiescent long-lived memory cells that remain poised to reproliferate following subsequent immunological challenges. Here we addressed whether T cells are intrinsically constrained by chronological or cell-division limits. We

activated mouse T cells in vivo using acute heterologous prime–boost–boost vaccinations[4], transferred expanded cells to new mice, and then repeated this process iteratively. Over 10 years (greatly exceeding the mouse lifespan)[5] and 51 successive immunizations, T cells remained competent to respond to vaccination. Cells required sufficient rest between stimulation events. Despite demonstrating the potential to expand the starting population at least 10^{40}-fold, cells did not show loss of proliferation control and results were not due to contamination with young cells. Persistent stimulation by chronic infections or cancer can cause T cell proliferative senescence, functional exhaustion and death[6]. We found that although iterative acute stimulations also induced sustained expression and epigenetic remodelling of common exhaustion markers (including PD1, which is also known as PDCD1, and TOX) in the cells, they could still proliferate, execute antimicrobial functions and form quiescent memory cells. These observations provide a model to better understand memory cell differentiation, exhaustion, cancer and ageing, and show that functionally competent T cells can retain the potential for extraordinary population expansion and longevity well beyond their organismal lifespan.

Main

The extent to which somatic mammalian cells can proliferate has been debated. Early reports that fetal chicken heart cells could grow in vitro for decades indicated that although organisms age, cells were immortal[7,8]. This sensationalistic experiment was not reproducible and attributed to cell transformation or continuous introduction of fresh cells[9]. The authors of ref. [10] putatively rejected diploid somatic cell immortality, reporting a division counter limiting human cells to 50–60 divisions. This is referred to as a ‘Hayflick limit’ and was ascribed to telomere shortening[3,11], proposed to reduce cancer but impose lifespan constraints, and species-specific in vitro cell division limits have been correlated with longevity (mouse cells undergoing fewer divisions and tortoise cells undergoing more)[2,12,13]. Species-specific lifespan limits may be intrinsic to other fundamental biological processes, including replication-dependent DNA mutations. For example, mice accumulate a similar number of DNA mutations per cell over their 3-year lifespan to that of humans over their 80-year lifespan, perhaps

revealing a conserved lifespan-defining boundary[14]. Other impediments to cell longevity or proliferative potential include loss of proteostasis and metabolic fitness.

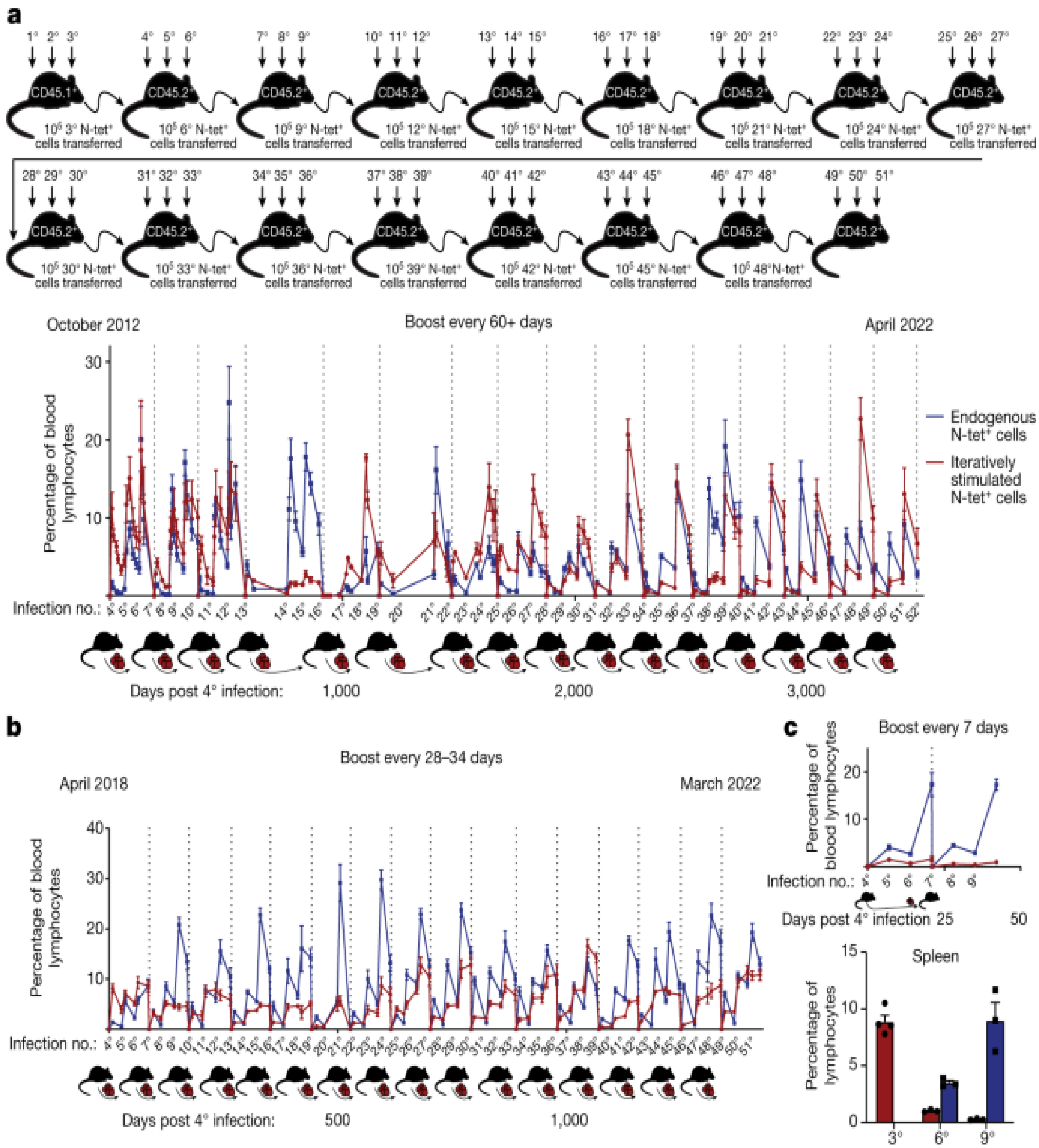
$CD8^+$ T lymphocytes can sustain quiescence for years, yet engage in a differentiation programme within minutes of activation, followed by accumulation of an unusual number of rapid cell divisions. Primary activated cells undergo a programmed burst of about 15–20 cell divisions, with the capacity for at least 3 cell divisions per day[15]. This results in massive clonal expansion and is the basis for the specificity and memory that typifies adaptive immunity. Whether cumulative T cell division potential is intrinsically finite or unlimited is relevant to control of chronic or recurring infections, cancer, vaccine boosting, age-associated immunologic senescence and adoptive cell therapies, as well as fundamental questions regarding mammalian cell biology.

Since the 1970s, there have been reports that T cells could occasionally be sustained in culture[16,17], although it has been debated whether this was the result of transformation[18], and the relevance of in vitro proliferation limits to organismal biology has been questioned[19,20]. Early attempts to grow anticancer T cells in vitro resulted in massive populations, but they survived poorly on transfer to patients[21,22,23]. This has been attributed to a lack of durable 'stemness' by T cells[24,25]. Antigen-experienced T cells that retain the ability to give rise to heterogenous daughter cells have been referred to as stem-like and provide more durable and expansible engraftment on transfer[26]. 'Memory stem cells' are typically described as sharing markers with naive T cells, including CD62L, although the property of unlimited division potential has not been tested[27]. Indeed, endogenous T cells responding to chronic infections or cancer undergo progressive in vivo 'exhaustion', defined by impaired proliferative ability, loss of potential to express effector cytokines and failure to eliminate immune targets[6]. When resting memory T cells have been serially passaged and stimulated with transient infections in vivo, they failed to respond to further stimulation within 4–7 generations[4,28,29]. These data support a model in which T cells have an intrinsic limit to their proliferation capacity, whether the antigen stimulation is chronic or periodic; however, this result may have depended on specific experimental conditions[30].

Proliferation and longevity potential

We wished to test whether: senescence, exhaustion and death are inevitable consequences of cumulative stimulations, divisions or species lifespan constraints; or T cells have the intrinsic potential to expand indefinitely. We designed an experiment that could isolate the potentially different effects of chronic versus cumulative stimulation, supersede species lifespan limits and permit iteratively potent $CD8^+$ T cell boosts by overcoming neutralizing antibody interference and competition for antigen by abundant memory T cell populations. $CD45.1^+$ C57Bl/6 female mice were immunized with three heterologous prime–boost immunizations (vesicular stomatitis virus subtype New Jersey (VSVnj); vaccinia virus expressing the VSV Indiana (VSVind) nucleoprotein (VVn); and VSVind (see Methods)) at ≥60-day intervals, which results in an abundantly expanded and long-lived 3° memory $CD8^+$ T cell population specific for the '$N_{52–59}$' peptide of VSV[30]. H-$2K^b$/$N_{52–59}$ major histocompatibility complex I $tetramer^+CD8^+$ T cells were then sorted from spleen and lymph nodes and 1×10^5 were transferred to congenic CD45.2 recipients, which subsequently received three additional immunizations (4° VSVind, 5° VVn and 6° VSVnj) at ≥60-day intervals. Although C57Bl/6 female mice survive only up to about 1,100 days[5], we were able to continue this sorting–transfer–three-immunization process for about 10 years totalling 51 cumulative immunization events because the $CD45.1^+$ population never lost the ability to expand following immunization (Fig. 1a). These cells also populated nonlymphoid tissues (Extended Data Fig. 1). These data showed that T cell expansion potential is not limited by the age of the origin population or the number of stimulation events.

Fig. 1: $CD8^+$ T cells can undergo seemingly unlimited bursts of proliferation if rested between stimulations.



a, CD45.1$^+$ VSV/N$_{52–59}$-specific memory CD8$^+$ T cells were iteratively sorted, transferred to CD45.2$^+$ mice, and then boosted three times per mouse with 60+ days between each of three heterologous immunizations. The graph shows the percentage of total lymphocytes in blood that are comprised of transferred CD45.1$^+$ iteratively boosted cells (red) or newly generated CD45.2$^+$ recipient H-2K^b/N$_{52–59}$-specific cells (blue). N-tet$^+$ = H-2K^b/N$_{52–59}$-tetramer. **b**,**c**, As in **a**, except boosting was carried out at 28–34-day (**b**) or

7-day (**c**) intervals. Similar results were observed in trailing cohorts that have undergone 13, 21 or 31 (≥60-day interval), or 15 stimulations (28–34-day interval). **c** is representative of two experiments with similar results with $n = 4$ at 3° and $n = 3$ at 6° and 9°. Error bars show average and s.e.m. For flow cytometry gating strategies, see Supplementary Fig. 1. For exact number of mice at each time point in **a**,**b**, see Supplementary Table 1.

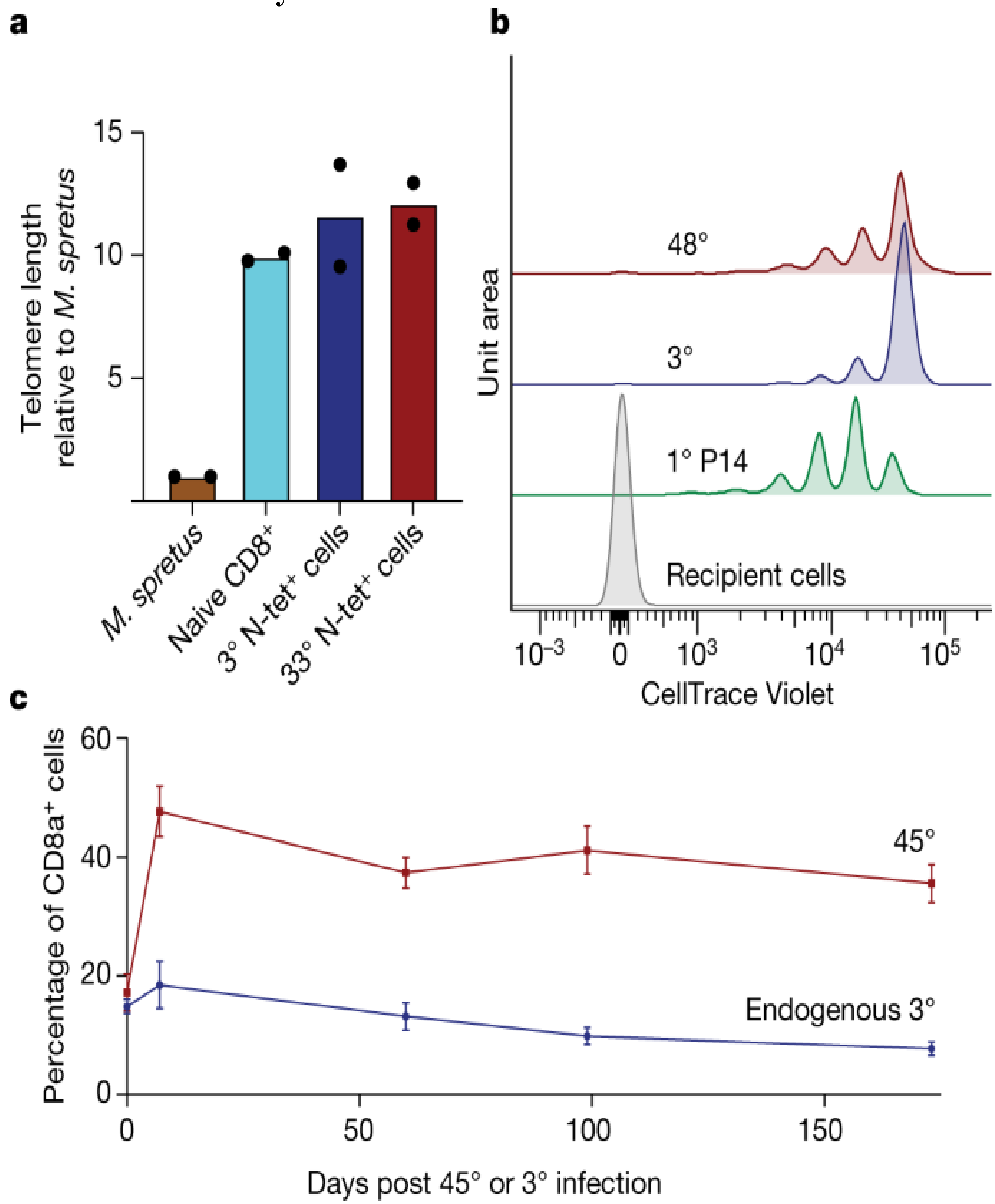
Source data

We then tested whether an interval of ≥60 days was required between boosts[31]. Vaccinations, cell sorting and transfers were carried out as in Fig. 1a, except with 28–34 days between each immunization. We observed no substantive loss of continued expansibility through 51 stimulations (Fig. 1b). However, boosting every 7 days resulted in rapid deterioration of expansion potential (Fig. 1c). Collectively, these data indicate that although iterative stimulation can result in senescence or death among T cells, this fate is not a biological imperative.

Cell cycle integrity

It has been proposed that cell proliferation is accompanied by telomere shortening that induces cellular senescence, prevents cancer and limits lifespan. Some have questioned this paradigm[19,32,33], and many dividing cells, including activated T cells, express telomerase[34,35,36]. As we observed >50 bursts of substantial proliferation among a T cell population that has outlived its species, we examined telomere length by quantitative PCR[37]. Telomere length was maintained (Fig. 2a). Iterative transfers would seemingly select for cancer cells that outgrow untransformed cells. We loaded iteratively stimulated T cells (ISTCs) with a cell-tracking dye, and then transferred them to mice without subsequent stimulation. Thirty-four days later, we found no evidence that ISTCs lost growth control; indeed, they underwent less homeostatic turnover than primary memory cells (Fig. 2b). Moreover, although the ISTC population was durable, it was not inflationary (Fig. 2c). Thus, memory T cells can accumulate many divisions over many years without malignant transformation or loss of durability.

Fig. 2: Iteratively boosted T cells maintain telomere length, cell cycle control and durability.



a, Naive, 3° memory or 33° memory $CD8^+$ T cells were sorted, and then tested for telomere length by quantitative PCR and compared to *Mus spretus* reference DNA. **b**, 3° and 48° cells were labelled with CellTrace Violet

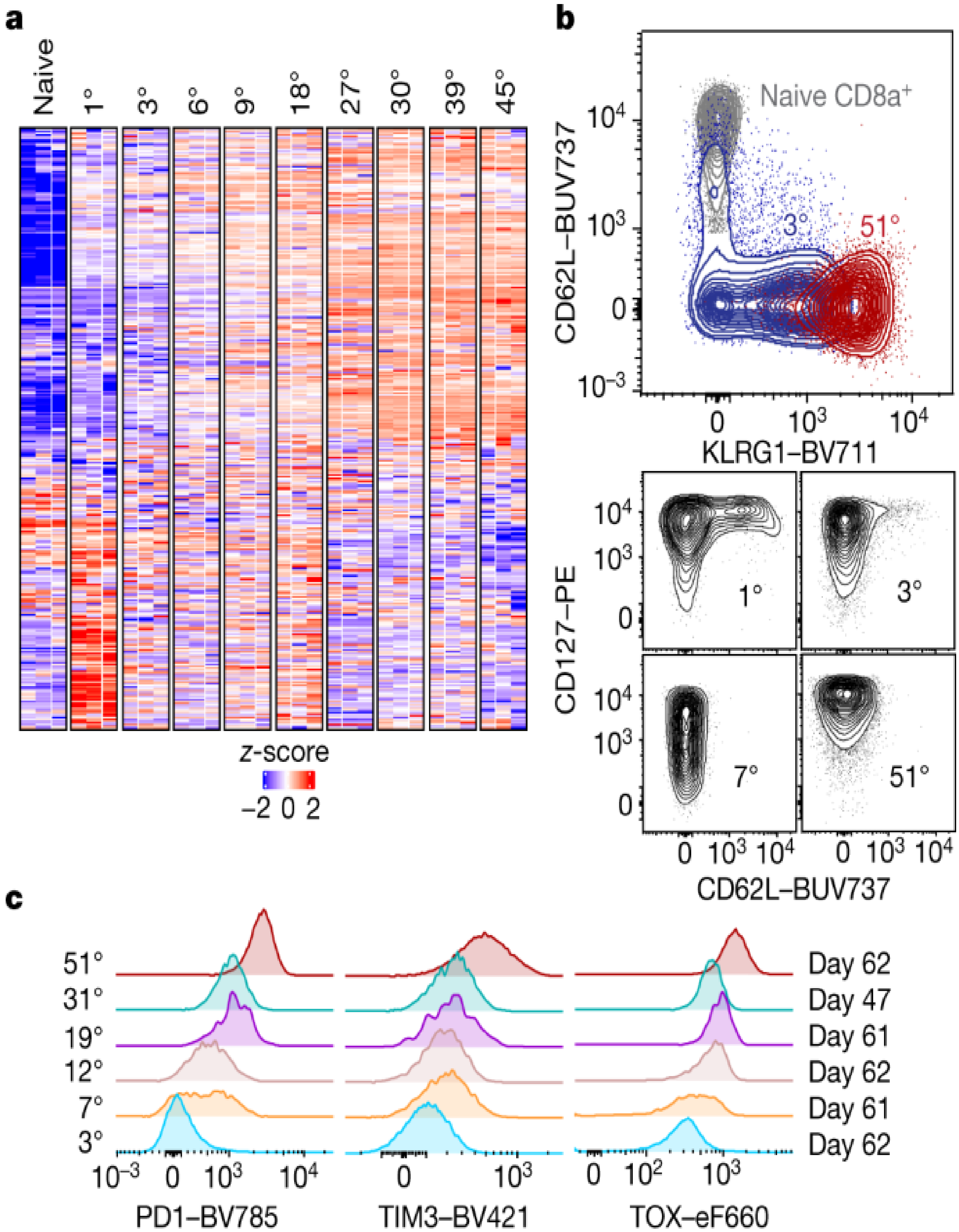
division-tracking dye, and then transferred to recipients without further boosting. Primary lymphocytic choriomeningitis virus (LCMV)-specific P14 memory $CD8^+$ T cells were transferred for comparison. Cumulative cell divisions, indicated by dye dilution, were evaluated in spleen 34 days later. **c**, 45° and endogenous 3° memory $CD8^+$ T cells were tracked in blood for 173 days following infection. **a–c** are representative of two experiments with similar results, $n = 2$ (**a**), $n = 4$ (**b**), $n = 5$ (**c**). Error bars show average and s.e.m. For flow cytometry gating strategies, see Supplementary Fig. 2.

Source data

Phenotypic evolution

Early claims of immortal propagation of untransformed cells were later attributed to continuous seeding with young cells[9]. We addressed this possibility by sorting ISTCs before every transfer on a genetically encoded marker combination unique to the primary immunized mouse ($CD45.1^+CD45.2^−$; Fig. 1a). Furthermore, boosting was associated with progressive changes in gene expression, ruling out contamination with a dominant young T cell population (Fig. 3a and Extended Data Fig. 2). Changes were also observed at the protein level. ISTCs did not resemble phenotypes associated with central or stem memory T cells that preponderate primary immune responses as they lacked CD62L and expressed the 'senescence marker' KLRG1 (Fig. 3b). Moreover, each generation progressively acquired molecules thought to define T cell exhaustion, including PD1, TIM3 (also known as HAVCR2) and TOX (Fig. 3c).

Fig. 3: Iterative boosting induces progressive changes in gene and protein expression and acquisition of exhaustion markers.



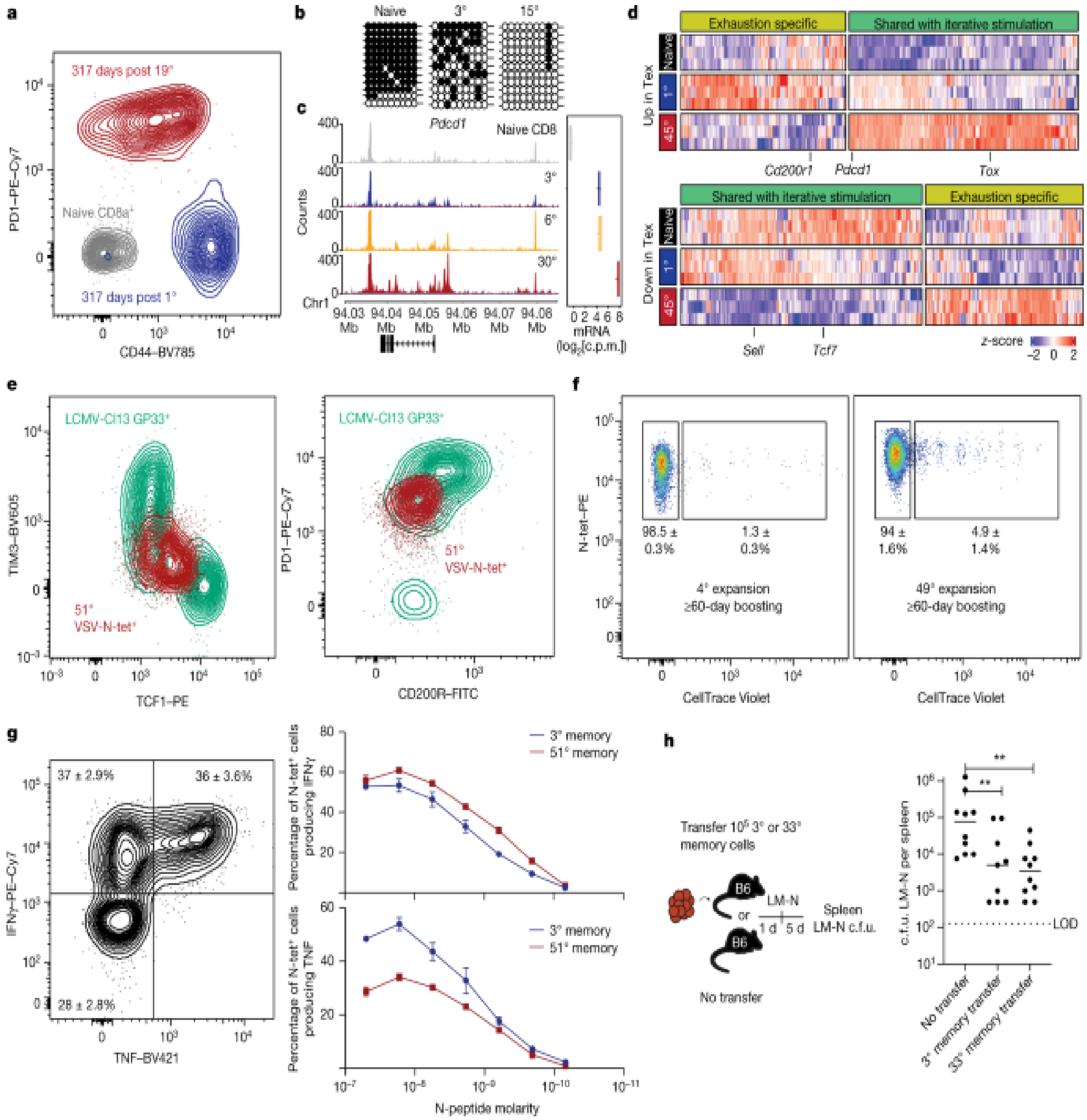
a, RNA-seq was carried out on naive and N-specific memory CD8^{+} T cells that had experienced progressive boosts. Shown is a heatmap of the genes that were previously reported[42] to be uniquely expressed by exhausted CD8^{+}

T cells. **b**,**c**, The phenotype of naive and various generations of H-2K^{b}/N$_{52–59}$-specific memory cells isolated from blood was assessed by flow cytometry. All samples were run on the same day from staggered cohorts. The right column in **c** indicates days after last boost. **a**, $n = 3$ per group. **b**,**c**, Representative of $n = 3$ (7°), $n = 4$ (3°, 12°, 19° and 31°) or $n = 9$ (51° and naive), from at least two independent experiments with similar results. For flow cytometry gating strategies, see Supplementary Fig. 3.

Maintenance of immunological functions

PD1 is expressed in response to recent T cell receptor stimulation and is maintained by chronic antigen stimulation[38]. As antigen is not known to persist after this vaccination regimen, we tested whether PD1 expression was epigenetically maintained after iterative acute boosts. When ISTCs rested in recipient mice for >300 days after a single boost, they still maintained PD1 whereas endogenous primary memory CD8^{+} T cells specific for the same antigen within the same mouse lacked PD1 (Fig. 4a). We also observed that CpG dinucleotides within the *Pdcd1* locus had become demethylated and that chromatin had become accessible (Fig. 4b,c). These data indicated that ISTCs phenocopied exhausted T cells exposed to chronic antigen, but the underlying regulation may be different. RNA sequencing (RNA-seq) and flow cytometric analyses revealed that ISTCs only partly acquire the exhaustion programme described for chronic viral infection in mice (Fig. 4d,e). ISTCs differed from exhausted T cells because they maintained the capacity to proliferate in response to antigen (Fig. 1). Indeed, when we transferred CellTrace Violet-labelled 3° and 48° ISTCs, a similar proportion of cells underwent ≥8 divisions following a single additional boost (Fig. 4f). Moreover, ISTCs did not lose the ability to produce interferon-γ (IFNγ) and substantial tumour necrosis factor (TNF) effector cytokines within 4–5 h of ex vivo peptide stimulation (Fig. 4g), nor were they diminished in the ability to control a microbial infection (Fig. 4h).

Fig. 4: Transcriptional, epigenetic and functional profiling distinguishes ISTCs from exhausted T cells.



a, 18° CD45.1^{+} ISTCs were transferred to naive mice before a single immunization with VSVind. 317 days later, ISTCs and endogenous VSV/N$_{52–59}$-specific primary memory and naive CD8^{+} T cells were assessed for maintenance of PD1 expression in blood. **b**,**c**, Methylation (**b**) or chromatin accessibility (**c**) of the *Pdcd1* locus (encoding PD1) was evaluated in naive, 3° or ISTC cells by bisulfite sequencing or assay for transposase-accessible chromatin using sequencing (ATAC-seq), respectively. *Pdcd1* RNA as measured by RNA-seq is also shown. Mb, megabase. **d**, Gene expression modules that are shared or unshared with genes reported to be upregulated or downregulated by exhausted CD8^{+} T

cells (Tex)[42]. **e**, Spleens with GP33-specific cells induced by chronic LCMV clone 13 (Cl13) or ISTCs were analysed by flow cytometry. **f**, 3° or 48° cells were labelled with CellTrace Violet division-tracking dye and transferred to recipients that were then infected with VSVind. The plots show cell division indicated by CellTrace Violet dilution 16 days after booster immunization. **g**, IFNγ and TNF staining 4 h after peptide stimulation. Representative flow cytometry (left panel) and response to peptide titration (right graphs). **h**, ISTCs or endogenous VSV/$N_{52–59}$-specific 3° memory cells were transferred to naive mice, which were then infected intravenously with *Listeria monocytogenes* expressing VSV-N (LM-N), and then assessed five days (5 d) later for bacterial burden in spleen. c.f.u., colony-forming units; LOD, limit of detection. $n = 3$ (**a**,**c**,**d**,**g**), $n = 4$ (**b**,**e**), $n = 4$ (4°) or 5 (49°) (**f**), $n = 9$ (3°) or 10 (no transfer and 33°) (**h**). **a**,**e**–**g** are representative of ≥2 similar experiments, **b**–**d** show data from a single experiment. **g** shows data combined from two experiments. Ordinary one-way analysis of variance with Dunnett's multiple-comparison test was used on $\log_{10}$-transformed data to test for significance between groups; $**P = 0.0086$ (for 3°) or $P = 0.0028$ (for 33°). Error bars show average and s.e.m. For flow cytometry gating strategies, see Supplementary Fig. 4. See Supplementary Information 3 for full statistical test details.

Source data

Discussion

These data demonstrate that T cells intrinsically have the capacity for seemingly unlimited population expansion and substantially outliving their host organism. Whether other somatic cells also exhibit this potential is unknown. Unexpectedly, we failed to capture ISTCs that detectably lost proliferation control, raising questions of whether antigen-experienced T cells are particularly poised to trigger cell death in response to pre-transformation events. This biology may accommodate the unusual lifestyle of T cells, characterized by long periods of quiescence punctuated by rapid bursts of extensive proliferation, and then programmed death of most of the expanded population. Perhaps ISTCs could help inform the biology of protection from cancer as well as maintaining cellular fitness far beyond the perceived constraints of time and experience.

Nevertheless, ISTCs changed epigenetically, transcriptionally and phenotypically with time and stimulation history. They lost canonical markers associated with memory T cell 'stemness'[27] and acquired markers associated with T cell exhaustion and dysfunction[6] (Fig. 3). Although they do not seem to expand equivalently to naive or central memory T cells on a per cell basis (Fig. 1a), progressive phenotypic changes in ISTCs were not parallelled by any detectable cumulative erosion in abilities to proliferate or form durable memory over the final 48 boosts (Fig. 2c). Our results do not reject the concept of T cell stemness, but dissociate self-renewing potential from markers previously associated with the ability to maintain longevity and proliferative fitness. ISTCs also dissociate exhaustion, as defined by proliferative ability, function and antigen-independent longevity, from phenotypes common to exhausted cells characterized in chronic infection or cancer. Functional $PD1^+$ T cells have also been observed in humans, and these may represent cells that have undergone iterative stimulation through heterosubtypic reinfections or pathogen recrudescence[39,40].

T cell senescence, exhaustion or death is frequently observed in settings of chronic stimulation, extremes of age and certain models of iterative stimulation punctuated by intervening rest[4,6,28,29,30,41]. Our results indicate that these fates are avoidable. Indeed, because ISTCs were diluted before every three boosts, a conservative estimate indicates that, on average, each of the approximately 200 $H\text{-}2K^b/N_{52–59}$-specific naive $CD8^+$ T cells present within unimmunized mice had the proliferative potential to give rise to more than 10^{41} 51° resting memory $CD8^+$ T cells (Extended Data Fig. 3), with a total cell volume of more than 30,000 times that of planet Earth. ISTCs provide a model to dissect the parameters that permit everlasting proliferative capacity and longevity among T cells, as well as the broader underlying biological mechanisms of maintaining fitness.

Methods

Mice

Donor female B6.SJL-Ptprc^{a}Pepc^b/BoyJ ($CD45.1^+$ B6) and P14 $CD8^+$ T cell transgenic mice were bred at the University of Minnesota animal facilities.

Female C57BL/6J (CD45.2$^+$ B6) mice were purchased from Jackson Laboratories and served as recipient mice, which were 8–10 weeks old at the time of first infection. Animals were housed with cycles of 14 h of light and 10 h of dark. Their environment was maintained at temperatures between 68 and 72 °F and humidity levels of 30–70%. Animals were treated according to the Institutional Animal Care and Use Committee guidelines and the protocols were approved by the Institutional Animal Care and Use Committee at the University of Minnesota.

Viral infections

CD45.1$^+$ tertiary memory cells were expanded through heterologous prime–boost–boost infection with 10^6 plaque-forming units (p.f.u.) VSVnj, an experiment-specific period of rest, 2×10^6 p.f.u. VVn, an experiment-specific period of rest, and 10^7 p.f.u. of VSVind. Following transfer to recipient CD45.2$^+$ mice, cells were expanded through heterologous prime–boost–boost infection with 10^6 p.f.u. of VSVind, an experiment-specific period of rest, 2×10^6 p.f.u. VVn, an experiment-specific period of rest, and 10^7 p.f.u. VSVnj. All heterologous prime–boost–boost infections were delivered through the tail vein. For LCMV Armstrong infections, 2×10^5 p.f.u. was delivered through an intraperitoneal injection. For LCMV clone 13, mice were given 200 μg of anti-mouse CD4 (GK1.5) from BioXCell through an intraperitoneal injection one day before and one day after an infection through the tail vein with 2×10^6 p.f.u. of virus.

Tracking of ISTCs

At various time points post infections, blood was collected from the submandibular vein. Red blood cells were lysed using ACK lysis buffer, and blood cells were stained with a variety of antibodies that always included anti-mouse CD8a (53-6.7; 1:100) from BD Biosciences, anti-mouse CD44 (IM7; 1:200) from BioLegend, anti-mouse CD45.1 (A20; 1:400), Ghost Dye Red 780 (1:1,000) from Tonbo Biosciences, and major histocompatibility complex I tetramer with the H2Kb-binding RGYVYQGL peptide from VSVind nucleoprotein (N-tetramer; 1:200). Major histocompatibility complex tetramers were prepared as previously described[43]. Flow cytometry

data were collected on a BD LSR II, BD Fortessa or Cytek Aurora and analysed using BD FlowJo. After a cell transfer, values for ISTCs were calculated on the basis of the number of cells transferred and 10% survival whereas values for endogenous cells were based on the reported number of naive $H2K^b$/RGYVYQGL-binding cells in B6 mice[44].

Fluorescence-activated cell sorting and cell transfers

A single-cell suspension was prepared from the spleen and macroscopic lymph nodes of donor mice. In some cases, red blood cells were lysed with ACK lysis buffer before the samples were stained with surface antibodies. In other cases, $CD8^+$ T cells were enriched through negative selection before they were stained with surface antibodies. $CD8^+$ T cell enrichment was carried out using the Stem Cell EasySep Mouse $CD8^+$ T Cell Isolation Kit following the manufacturer's instructions or using a prepared cocktail of biotinylated antibodies. Briefly, single-cell suspensions were resuspended at 10^8 cells per millilitre in phosphate-buffered saline (Gibco) supplemented with 2% heat-inactivated fetal bovine serum (Peak Serum) and 1 mM EDTA (Promega) and then incubated with 5% rat serum (Stem Cell) and 0.0275 mg ml^{-1} anti-mouse CD4 (GK1.5) from Invitrogen and anti-mouse CD19 (1D3), anti-mouse CD11b (M1/70), anti-mouse NK1.1 (PK136), anti-mouse F4/80 (BM8.1), anti-mouse TER119, anti-mouse CD45R (RA3-6B2), anti-mouse LY6G (GR1) and anti-mouse CD16/32 (2.4G2) from Tonbo Biosciences. All antibodies were conjugated to biotin. After antibody incubation, antibody-bound cells were removed using the Stem Cell EasySep Mouse Streptavidin RaphidSpheres Isolation kit following the manufacturer's instructions. Cells were then stained with anti-mouse CD8a (53-6.7; 1:200) from BD Biosciences, anti-mouse CD45.1 (A20; 1:200), anti-mouse CD45.2 (104; 1:200), N-tetramer (1:200) and Ghost Dye Red 780 (1:1,000) from Tonbo Biosciences. Live $CD8a^+$VSV-N-$tetramer^+CD45.1^+CD45.2^-$ cells were sorted on a BD FACS Aria II, and 10^5 sorted cells were transferred through the tail vein into recipient mice. Infections resumed the following day.

Quantitative-PCR-based measurement of mouse telomere length

Telomere length was measured using quantitative PCR, including the primers (synthesized by Integrated DNA Technologies) and control gene as described previously[37]. Data were collected using the QuantStudio 5 system (Applied Biosystems). Isolated *M. spretus* DNA was purchased from Jackson Laboratories to serve as a source of mouse DNA with relatively short telomeres.

CellTrace Violet labelling of cells

The cells of a single-cell suspension of cells isolated from the spleen and macroscopic lymph nodes were labelled with CellTrace Violet (Invitrogen) following the manufacturer's instructions. CellTrace Violet-labelled cells were transferred into mice through the tail vein.

RNA-seq

Bulk RNA was isolated from 10^5 sorted cells using the Qiagen RNeasy Plus Micro kit following the manufacturer's instructions. Libraries were prepared using the Takara/Clontech Stranded Total RNA-seq pico input mammalian kit. Naive, 1°, 3° and 27° cells were prepared using kit version 1 and all other samples were prepared used kit version 2. Samples prepared with kit version 1 were sequenced on an Illumina HiSeq; samples prepared with kit version 2 were sequenced on the Illumina NovaSeq 6000. Quality of fastq files was assessed with FastQC. Adapters and low-quality segments were trimmed with Trimmomatic. Filtered reads were aligned to the mouse genome GRCm38 using Hisat2 and the count matrix was generated using featureCounts. All subsequent gene expression data analyses were carried out in the R software. Genes expressed at low levels were filtered using the filterByExpr function and TMM-normalized in edgeR. Differentially expressed genes were determined using limma. The RNA-seq samples were sequenced in two batches. Genes were considered significantly different if log[fold change] > 1 and false discovery rate < 0.05. Heatmaps were generated with the ComplexHeatmap package. We noticed a batch effect between the first and the second run of RNA-seq samples. However, differential gene expression analysis between the two naive groups revealed only few differentially expressed genes, most being undefined genes or

ribosomal genes, indicating that the observed batch effect is not confounding our analysis.

Cell phenotyping through flow cytometry

Phenotyping was carried out on either blood cells or splenocytes that were treated with ACK lysis buffer. Cells were stained extracellularly with various combinations of: anti-mouse CD8a (53-6.7; 1:100), anti-mouse CD4 (GK1.5; 1:1,000), anti-mouse CD45.1 (A20; 1:400), anti-mouse CD122 (TM-β1; 1:100), anti-mouse CD62L (MEL-14; 1:800) and anti-mouse KLRG1 (2F1; 1:200) from BD Biosciences, anti-mouse CD44 (IM7; 1:200), anti-mouse CD38 (90; 1:100), anti-mouse CD45.1 (A20; 1:400), anti-mouse CD28 (E18; 1:100), anti-mouse PD1 (RMP1-30; 1:100), anti-mouse CD200R (OX-110; 1:50) and anti-mouse TIM3 (RMT3-23; 1:100) from BioLegend, anti-mouse CD45.1 (A20; 1:400), anti-mouse CD127 (A7R34) (1:100), anti-mouse CD45.2 (104; 1:200) and Ghost Dye Red 780 (1:1,000) from Tonbo Biosciences and either N-tetramer or the H2Db-binding KAVYNFATM peptide from lymphocytic choriomeningitis virus (GP33-tetramer), fixed and permeabilized using Tonbo Foxp3/Transcription Factor Staining Kit, and then stained intracellularly in Tonbo Permeabilization buffer with anti-mouse TOX (TXRX10; 1:50), and anti-mouse EOMES (Dan11mag; 1:50) from Invitrogen, anti-mouse BCL-2 (10C4; 1:50) from eBioscience, and anti-mouse TCF1/TCF7 (C63D9; 1:50) from Cell Signaling Technology. Flow cytometry data were collected on either a BD Fortessa or Cytek Aurora and analysed with BD FlowJo.

ATAC-seq

ATAC-seq was carried out following a previously described protocol[45]. Library preparation on transposed DNA was carried out with a Nextera DNA library preparation kit following the manufacturer's instructions. Samples were sequenced on an Illumina HiSeq. FastQC was used to assess the quality of fastq files. Reads were aligned to the mouse Genome (UCSC version mm10) using bowtie2 with valid alignment per read and allowed numbers of mismatches set to 1. Samtools was used to generate sorted bam files that contain mapped reads alone, and Picard was used to mark duplicates. All subsequent data analyses were carried out in the R software.

Blacklist regions and mitochondrial reads were removed before peak regions were called with csaw using a window width of 200 base pairs. Windows were considered enriched over background if the $\log_2$[fold change] was >3. As biological sample groups strongly differed in the enrichment of reads within accessible chromatin over background, log[c.p.m.] TMM-scaled peak counts were quantile-normalized as previously described[46]. Differential accessibility analysis was carried out with limma, and peak regions were subsequently merged with a maximum distance between adjacent windows of 200 base pairs. Genomic regions were plotted using Gviz.

PD1 methylation analysis

Genomic DNA was isolated from the purified cells and subjected to bisulfite treatment using the Zymo EZ DNA methylation kit as per the manufacturer's instructions. Bisulfite-modified DNA was PCR-amplified using PD1 promoter-specific primers[47]. The amplicon was cloned into TA vector and transformed into bacteria. Vector from individual colonies was sequenced and analysed using the BISMA software (Bremen, Germany), as previously described[47].

In vitro cytokine stimulation

Blood cells were collected from the submandibular vein, red blood cells were lysed using ACK lysis buffer, and cells were resuspended in RPMI supplemented with 5% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U ml^{-1} penicillin–streptomycin and 0.05 mM β-mercaptoethanol. Cells were added to the wells of a 96-well plate and medium containing brefeldin A from Tonbo Biosciences and various concentrations of RGYVYQGL VSV-N peptide synthesized by New England Peptide Inc (now named Vivitide) to yield a final concentration of 3 μM brefeldin A and labelled concentrations of peptide. After 4–5 h, cells were stained extracellularly with anti-mouse CD8a (53-6.7; 1:100) from BD Biosciences, anti-mouse CD45.1 (A20; 1:400), anti-mouse CD45.2 (104; 1:400) and Ghost Dye Red 780 (1:1,000) from Tonbo Biosciences and N-tetramer (1:200), fixed and permeabilized using Tonbo Foxp3/Transcription Factor Staining Kit, and then stained intracellularly in Tonbo Permeabilization buffer with TNF (MP6-XT22; 1:100) from BD Biosciences

and IFNγ (XMG1.2; 1:100) from BioLegend. Flow cytometry data were collected on a BD Fortessa and analysed with BD FlowJo software.

Listeria protection

One day after transfer of 10^5 specified sorted N-specific cells, mice were injected through the tail vein with 7×10^3 colony-forming units of *L. monocytogenes* expressing VSV-N (LM-N). Five days after LM-N infection, spleens were removed and cells were lysed through homogenization in sterile 0.5% Igepal CA-630 (Sigma Aldrich). Various dilutions of cell homogenate were plated onto Petri dishes with BBL Brain Heart Infusion Agar (BD Biosciences) prepared according to the manufacturer's instructions, incubated at 37 °C overnight and colonies were counted the next day.

Non-lymphoid tissue analysis

Non-lymphoid tissues were treated as previously described[48], including the use of 3 µg of an intravascular anti-CD8a antibody (53-6.7) from BD Biosciences as previously described[49] to prepare a single-cell suspension. Single-cell suspensions from each tissue were stained extracellularly with anti-mouse CD8a (53-6; 1:00) from BD Biosciences, anti-mouse CD45.1 (A20; 1:200) and Ghost Dye Red 780 (1:1,000) from Tonbo Biosciences, anti-mouse CD69 (H1.2F3; 1:50) from Invitrogen and N-tetramer (1:200). Flow cytometry data were collected on a BD Fortessa and analysed with BD FlowJo software.

Data measurements

Cells that were tracked over time were measured repeatedly after different infections with discreet measurements made on populations of cells within different mice.

Statistical methods

No statistical test was used to determine sample sizes. Mice were randomly assigned to different experimental groups. Researchers were not blinded to

experimental groups. Specific statistical tests used to determine significance, group sizes (*n*) and *P* values are provided in the figure legends. P value < 0.05, significant. All statistical analysis was carried out using Prism (GraphPad).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Sequencing data are available through GEO (accession number GSE213230). Source data are provided with this paper.

References

1. Stanley, J. F., Pye, D. & MacGregor, A. Comparison of doubling numbers attained by cultured animal cells with life span of species. *Nature* **255**, 158–159 (1975).

2. Röhme, D. Evidence for a relationship between longevity of mammalian species and life spans of normal fibroblasts in vitro and erythrocytes in vivo. *Proc. Natl Acad. Sci. USA* **78**, 5009–5013 (1981).

3. Bodnar, A. G. et al. Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).

4. Masopust, D., Ha, S.-J., Vezys, V. & Ahmed, R. Stimulation history dictates memory CD8 T cell phenotype: implications for prime-boost vaccination. *J. Immunol.* **177**, 831–839 (2006).

5. Kunstyr, I. & Leuenberger, H. G. Gerontological data of C57BL/6J mice. I. Sex differences in survival curves. *J. Gerontol.* **30**, 157–162 (1975).

6. McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T cell exhaustion during chronic viral infection and cancer. *Annu. Rev. Immunol.* **37**, 457–495 (2019).

7. Carrel, A. On the permanent life of tissues outside of the organism. *J. Exp. Med.* **15**, 516–528 (1912).

8. Ebeling, A. H. A ten year old strain of fibroblasts. *J. Exp. Med.* **35**, 755–759 (1922).

9. Witkowski, J. A. Dr. Carrel's immortal cells. *Med. Hist.* **24**, 129–142 (1980).

10. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621 (1961).

11. Allsopp, R. C. et al. Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 10114–10118 (1992).

12. Goldstein, S. Aging in vitro growth of cultured cells from the Galapagos tortoise. *Exp. Cell Res.* **83**, 297–302 (1974).

13. de Magalhães, J. P. & Toussaint, O. Telomeres and telomerase: a modern fountain of youth? *Rejuv. Res.* **7**, 126–133 (2004).

14. Cagan, A. et al. Somatic mutation rates scale with lifespan across mammals. *Nature* **604**, 517–524 (2022).

15. Boer, R. J. D., Homann, D. & Perelson, A. S. Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *J. Immunol.* **171**, 3928–3935 (2003).

16. Morgan, D. A., Ruscetti, F. W. & Gallo, R. Selective in vitro growth of T lymphocytes from normal human bone marrows. *Science* **193**, 1007–1008 (1976).

17. Gillis, S. & Smith, K. A. Long term culture of tumour-specific cytotoxic T cells. *Nature* **268**, 154–156 (1977).

18. Effros, R. B., Dagarag, M., Spaulding, C. & Man, J. The role of CD8+ T-cell replicative senescence in human aging. *Immunol. Rev.* **205**, 147–157 (2005).

19. Rubin, H. The disparity between human cell senescence in vitro and lifelong replication in vivo. *Nat. Biotechnol.* **20**, 675–681 (2002).

20. Akbar, A. N. & Henson, S. M. Are senescence and exhaustion intertwined or unrelated processes that compromise immunity? *Nat. Rev. Immunol.* **11**, 289–295 (2011).

21. Rosenberg, S. A. et al. Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. *N. Engl. J. Med.* **319**, 1676–1680 (1988).

22. Rosenberg, S. A. et al. Treatment of patients with metastatic melanoma with autologous tumor-infiltrating lymphocytes and interleukin 2. *J. Natl Cancer Inst.* **86**, 1159–1166 (1994).

23. Dudley, M. E. & Rosenberg, S. A. Adoptive-cell-transfer therapy for the treatment of patients with cancer. *Nat. Rev. Cancer* **3**, 666–675 (2003).

24. Zhang, Y., Joe, G., Hexner, E., Zhu, J. & Emerson, S. G. Host-reactive CD8+ memory stem cells in graft-versus-host disease. *Nat. Med.* **11**, 1299–1305 (2005).

25. Gattinoni, L. et al. Wnt signaling arrests effector T cell differentiation and generates CD8+ memory stem cells. *Nat. Med.* **15**, 808–813 (2009).

26. Gattinoni, L. et al. A human memory T cell subset with stem cell–like properties. *Nat. Med.* **17**, 1290–1297 (2011).

27. Graef, P. et al. Serial transfer of single-cell-derived immunocompetence reveals stemness of CD8+ central memory T cells. *Immunity* **41**, 116–126 (2014).

28. Wirth, T. C. et al. Repetitive antigen stimulation induces stepwise transcriptome diversification but preserves a core signature of memory CD8+ T cell differentiation. *Immunity* **33**, 128–140 (2010).

29. Rai, D., Martin, M. D. & Badovinac, V. P. The longevity of memory CD8 T cell responses after repetitive antigen stimulations. *J. Immunol.* **192**, 5652–5659 (2014).

30. Fraser, K. A., Schenkel, J. M., Jameson, S. C., Vezys, V. & Masopust, D. Preexisting high frequencies of memory CD8+ T cells favor rapid memory differentiation and preservation of proliferative potential upon boosting. *Immunity* **39**, 171–183 (2013).

31. Thompson, E. A., Beura, L. K., Nelson, C. E., Anderson, K. G. & Vezys, V. Shortened intervals during heterologous boosting preserve memory CD8 T cell function but compromise longevity. *J. Immunol.* **196**, 3054–3063 (2016).

32. Rubin, H. Telomerase and cellular lifespan: ending the debate? *Nat. Biotechnol.* **16**, 396–397 (1998).

33. Mondello, C. et al. Telomere length in fibroblasts and blood cells from healthy centenarians. *Exp. Cell Res.* **248**, 234–242 (1999).

34. Weng, N. P., Levine, B. L., June, C. H. & Hodes, R. J. Regulated expression of telomerase activity in human T lymphocyte development and activation. *J. Exp. Med.* **183**, 2471–2479 (1996).

35. Weng, N., Hathcock, K. S. & Hodes, R. J. Regulation of telomere length and telomerase in T and B cells: a mechanism for maintaining replicative potential. *Immunity* **9**, 151–157 (1998).

36. Hathcock, K. S., Kaech, S. M., Ahmed, R. & Hodes, R. J. Induction of telomerase activity and maintenance of telomere length in virus-specific effector and memory CD8+ T cells. *J. Immunol.* **170**, 147–152 (2003).

37. Callicott, R. J. & Womack, J. E. Real-time PCR assay for measurement of mouse telomeres. *Comp. Med.* **56**, 17–22 (2006).

38. Bally, A. P. R., Austin, J. W. & Boss, J. M. Genetic and epigenetic regulation of PD-1 expression. *J. Immunol.* **196**, 2431–2437 (2016).

39. Duraiswamy, J. et al. Phenotype, function, and gene expression profiles of programmed death-1hi CD8 T cells in healthy human adults. *J. Immunol.* **186**, 4200–4212 (2011).

40. Sekine, T. et al. TOX is expressed by exhausted and polyfunctional human effector memory CD8+ T cells. *Sci. Immunol.* **5**, eaba7918 (2020).

41. Lee, K. et al. Characterization of age-associated exhausted CD8+ T cells defined by increased expression of Tim-3 and PD-1. *Aging Cell* **15**, 291–300 (2016).

42. Bengsch, B. et al. Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. *Immunity* **48**, 1029–1045 (2018).

43. Altman, J. D. & Davis, M. M. MHC-peptide tetramers to visualize antigen-specific T cells. *Curr. Protoc. Immunol.* **115**, 17.3.1–17.3.44 (2016).

44. Obar, J. J., Khanna, K. M. & Lefrançois, L. Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection. *Immunity* **28**, 859–869 (2008).

45. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).

46. Künzli, M. et al. Long-lived T follicular helper cells retain plasticity and help sustain humoral immunity. *Sci. Immunol.* **5**, eaay5552 (2020).

47. Youngblood, B. et al. Chronic virus infection enforces demethylation of the locus that encodes PD-1 in antigen-specific CD8+ T cells. *Immunity* **35**, 400–412 (2011).

48. Steinert, E. M. et al. Quantifying memory CD8 T cells reveals regionalization of immunosurveillance. *Cell* **161**, 737–749 (2015).

49. Anderson, K. G. et al. Intravascular staining for discrimination of vascular and tissue leukocytes. *Nat. Protoc.* **9**, 209–222 (2014).

Acknowledgements

We thank members of the laboratories of D.M. and V.V. for helpful discussions. We thank the University of Minnesota Flow Cytometry resource for cell sorting (J. Motl, T. Martin and R. Arora). This study was supported by National Institutes of Health grants R01 AI084913, R01 AI146032 (D.M.) and T32HL007741 (A.G.S.), and Swiss National Science Foundation grant P2BSP3_200187 (M.K.).

Author information

Authors and Affiliations

1. Center for Immunology, Department of Microbiology and Immunology, University of Minnesota, Minneapolis, MN, USA

   Andrew G. Soerens, Marco Künzli, Clare F. Quarnstrom, Milcah C. Scott, Lee Swanson, JJ. Locquiao, Vaiva Vezys & David Masopust

2. Department of Microbial Infection and Immunity, College of Medicine, The Ohio State University, Columbus, OH, USA

   Hazem E. Ghoneim

3. Division of Animal Physiology and Immunology, School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

   Dietmar Zehn

4. Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN, USA

Benjamin Youngblood

Contributions

A.G.S., M.K., C.F.Q., M.C.S., L.S., J.L. and H.E.G. carried out the experiments, analysed data and prepared visualizations. B.Y. analysed data and prepared visualizations. D.Z. created and shared reagents. A.G.S., M.K., V.V. and D.M. designed the experiments and wrote the manuscript.

Corresponding author

Correspondence to David Masopust.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Susan Kaech and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Iteratively stimulated T cells populate non-lymphoid tissues.

a) CD45.1+ 48° ISTC memory CD8 T cells were transferred to naïve mice, followed by three heterologous prime-boost-boost immunizations. Non-lymphoid tissues were analyzed 48 days after the 51° boost. Flow cytometry plots are gated on live CD8a+ lymphocytes that were not stained by intravascular in vivo antibody labeling. b) CD69 expression on endogenous 3° cells (CD45.1-) and 48° memory CD8 T cells (CD45.1+).Plots concatenated from four mice and experiment is representative of three similar experiments with similar results.

Extended Data Fig. 2 Gene set enrichment analysis of iteratively stimulated CD8 T cells.

Gene set enrichment analysis was performed on differentially expressed genes between 45° ISTCs and primary memory cells using the category "Biological process" of the Gene Ontology database. The top 20 upregulated (activated) and top 20 downregulated (suppressed) pathways in ISTCs are shown. Statistical significance was determined using Over-Representation analysis.

Extended Data Fig. 3 A single naïve CD8 T cell shows the potential to produce >10^{40} memory cell progeny.

a) Schematic showing estimate of cell expansion potential after 51 HPBB immunizations. HPBB: three successive Heterologous Prime, Boost, Boost immunizations. b) Schematic comparing the volume of earth with the theoretical volume of memory T cells implied by the calculated proliferation potential of a naïve CD8 T cell.

Supplementary information

Supplementary Information 1

This file contains Supplementary Figs. 1–4 showing the gating schemes used in Figs. 1–4 and Supplementary Table 1, which shows the number of animals at each time point in Fig. 1a,b.

Reporting Summary

Supplementary Information 2

This file lists the differentially expressed genes between 45° cells and 3° cells, 1° or naive $CD8^+$ T cells. Empirical Bayes moderation was used to test for significance.

Supplementary Information 3

Extended statistical results showing the full details of statistics carried out for Fig. 4g.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 4

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

About this article

Cite this article

Soerens, A.G., Künzli, M., Quarnstrom, C.F. *et al.* Functional T cells are capable of supernumerary cell division and longevity. *Nature* **614**, 762–766 (2023). https://doi.org/10.1038/s41586-022-05626-9

- Received: 01 July 2022
- Accepted: 05 December 2022
- Published: 18 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05626-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **You can’t live forever — but perhaps your T cells can**
  - Yvonne Bordon

  Nature Reviews Immunology (2023)

Immune cells’ ability to persist and replicate long outlives species lifespan

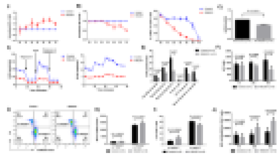
Nature Research Briefing 18 Jan 2023

T cells proliferate beyond organismal lifespan

- Anna Kriebs

Nature Aging Research Highlight 09 Feb 2023

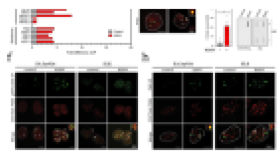
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05626-9



Telomeric injury by KML001 in human T cells induces mitochondrial dysfunction through the p53-PGC-1α pathway

02 December 2020

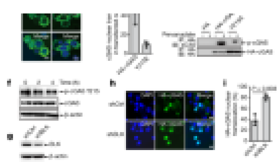
Madison Schank, Juan Zhao, … Zhi Q. Yao



Oncogenic herpesvirus KSHV triggers hallmarks of alternative lengthening of telomeres

21 January 2021

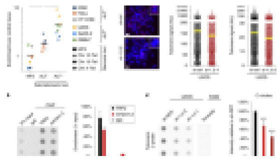
Timothy P. Lippert, Paulina Marzec, … Simon J. Boulton



Nuclear cGAS suppresses DNA repair and promotes tumorigenesis

24 October 2018

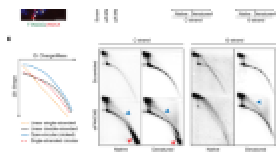
Haipeng Liu, Haiping Zhang, … Baoxue Ge



[The macroH2A1.2 histone variant links ATRX loss to alternative telomere lengthening]

04 March 2019

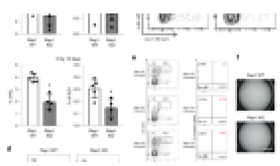
Jeongkyu Kim, Chongkui Sun, … Philipp Oberdoerffer



[The FANCM-BLM-TOP3A-RMI complex suppresses alternative lengthening of telomeres (ALT)]

28 May 2019

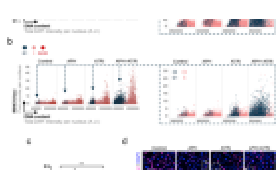
Robert Lu, Julienne J. O’Rourke, … Hilda A. Pickett



[Rap1 regulates hematopoietic stem cell survival and affects oncogenesis and response to chemotherapy]

13 December 2019

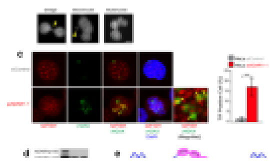
Ekta Khattar, Kyaw Ze Ya Maung, … Vinay Tergaonkar



[**RPA shields inherited DNA lesions for post-mitotic DNA synthesis**]

22 June 2021

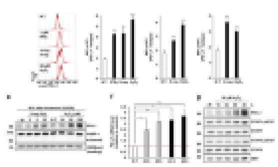
Aleksandra Lezaja, Andreas Panagopoulos, … Matthias Altmeyer



[**ADAR1 RNA editing enzyme regulates R-loop formation and genome stability at telomeres in cancer cells**]

12 March 2021

Yusuke Shiromoto, Masayuki Sakurai, … Kazuko Nishikura



[**Base excision repair regulates PD-L1 expression in cancer cells**]

12 February 2019

Tiara Bunga Mayang Permata, Yoshihiko Hagiwara, … Atsushi Shibata

- Article
- Open Access
- Published: 08 February 2023

Telomere-to-mitochondria signalling by ZBP1 mediates

replicative crisis

- Joe Nassour,
- Lucia Gutierrez Aguiar,
- Adriana Correia,
- Tobias T. Schmidt,
- Laura Mainz,
- Sara Przetocka,
- Candy Haggblom,
- Nimesha Tadepalle,
- April Williams,
- Maxim N. Shokhirev,
- Semih C. Akincilar,
- Vinay Tergaonkar,
- Gerald S. Shadel &
- …
- Jan Karlseder

Show authors

Nature **volume 614**, pages 767–773 (2023)

- 19k Accesses
- 254 Altmetric
- Metrics details

Abstract

Cancers arise through the accumulation of genetic and epigenetic alterations that enable cells to evade telomere-based proliferative barriers and achieve immortality. One such barrier is replicative crisis—an autophagy-dependent program that eliminates checkpoint-deficient cells with unstable telomeres and other cancer-relevant chromosomal aberrations[1,2]. However, little is known about the molecular events that regulate the onset of this important

tumour-suppressive barrier. Here we identified the innate immune sensor Z-DNA binding protein 1 (ZBP1) as a regulator of the crisis program. A crisis-associated isoform of ZBP1 is induced by the cGAS–STING DNA-sensing pathway, but reaches full activation only when associated with telomeric-repeat-containing RNA (TERRA) transcripts that are synthesized from dysfunctional telomeres. TERRA-bound ZBP1 oligomerizes into filaments on the outer mitochondrial membrane of a subset of mitochondria, where it activates the innate immune adapter protein mitochondrial antiviral-signalling protein (MAVS). We propose that these oligomerization properties of ZBP1 serve as a signal amplification mechanism, where few TERRA–ZBP1 interactions are sufficient to launch a detrimental MAVS-dependent interferon response. Our study reveals a mechanism for telomere-mediated tumour suppression, whereby dysfunctional telomeres activate innate immune responses through mitochondrial TERRA–ZBP1 complexes to eliminate cells destined for neoplastic transformation.

Main

Replicative senescence and crisis constitute two anti-proliferative barriers that human cells must evade to gain immortality[3]. Senescence is p53 and RB dependent and occurs when shortened telomeres elicit a DNA-damage response[4,5]. Loss of cell cycle checkpoints renders cells incapable of senescence activation, resulting in continued proliferation and telomere shortening. Such cells eventually succumb to crisis, characterized by extensive cell death and genome instability[1,2,6,7]. Crisis provides a redundant tumour-suppressor mechanism for replicative senescence, whereby CGAS–STING signalling triggers a non-canonical form of autophagy capable of executing cell death rather than sustaining cell survival[1]. Although this discovery established an essential role for autophagy during cell death in crisis, it remained unclear how dysfunctional telomeres engage nucleic-acid-sensing machineries and activate innate immune signalling pathways that are required for cell death.

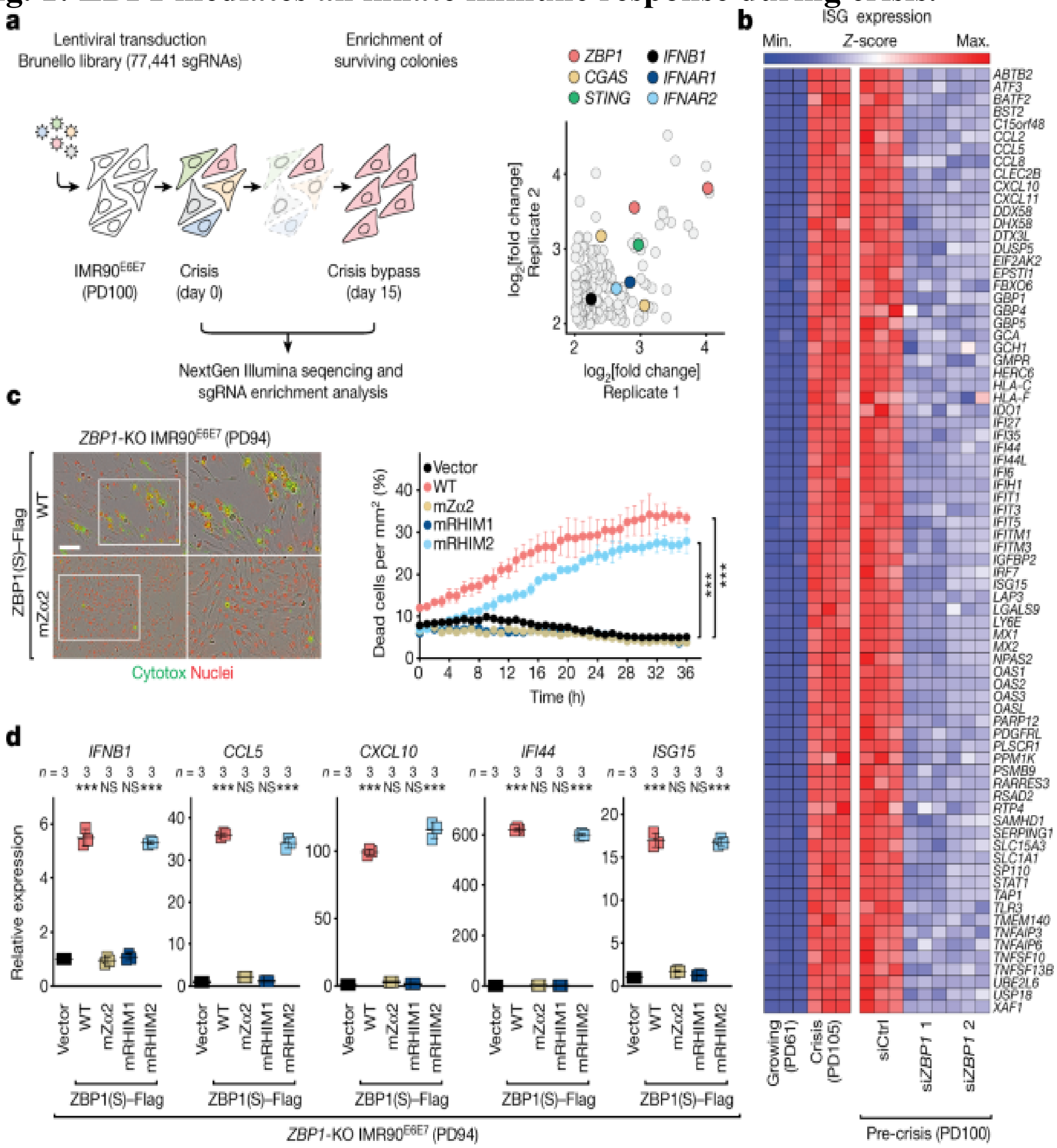
ZBP1 mediates an innate immune response in crisis

To study crisis, human papillomavirus E6 and E7 or SV40 large T antigen (SV40LT) were introduced into primary human IMR90 (IMR90^{E6E7}) and WI38 (WI38^{SV40LT}) lung fibroblasts[1,2], respectively, thereby silencing or disrupting the p53 and RB tumour-suppressor pathways[8,9,10,11]. Checkpoint-deficient cells bypassed senescence and reached crisis at population doubling 105–110 (PD105–110) for IMR90^{E6E7} and PD85–90 for WI38^{SV40LT} (refs. [1,2]). RNA sequencing (RNA-seq) analysis revealed profound transcriptional changes during crisis with an overlap of upregulated genes (Extended Data Fig. 1a,b). Crisis-associated processes were predominantly linked to innate immunity and inflammation (Extended Data Fig. 1c). Interferon (IFN)-stimulated genes (ISGs) (such as *CD74*, *DDIT4*, *GBP2*, *ISG20* and *RTP4*) were induced, consistent with an inflammatory status associated with replicative crisis (Extended Data Fig. 2a,b).

The inherent stringency of replicative crisis offered a powerful system in which to conduct a positive selection CRISPR–Cas9 knockout screen (Fig. 1a (left)). Survivors are expected to have lost pathways that are required for crisis, such as the ones linking dysfunctional telomeres to innate immune activation. Enrichment analysis revealed first that Gene Ontology (GO) terms associated with inflammation and innate immunity were predominant (Extended Data Fig. 2c). Second, the innate immune sensor ZBP1 (also known as DAI and DLM-1) emerged as a top hit, along with the previously characterized CGAS–STING–IFN pathway[1] (Fig. 1a (right)). ZBP1, an ISG product[12], was described as cytosolic nucleic acid sensor that induces type I IFNs and regulates innate immunity and cell death[13]. CRISPR-mediated deletion of *ZBP1* demonstrated its essential role in crisis. Control cells (sgLUC, sgGFP) entered crisis around PD90 for WI38^{SV40LT} and PD107 for IMR90^{E6E7}, when cell death was frequent and replicative ability was reduced (Extended Data Fig. 3a–d). However, pooled *ZBP1*-knockout cells continued to proliferate for an additional 7–10 population doublings beyond the crisis plateau with a notable reduction in cell death and maintenance of growth potential (Extended Data Fig. 3a–d). Cells that have bypassed crisis showed reduced expression and secretion of IFNβ, indicative of impaired type I IFN activity (Extended Data Fig. 3e,f). Depletion of ZBP1 abrogated the ISG signature, confirming that ISG induction during crisis is attributable

to ZBP1 (Fig. 1b). These results directly linked ZBP1 to innate immune activation and cell death during crisis.

Fig. 1: ZBP1 mediates an innate immune response during crisis.



a, Pre-crisis IMR90^{E6E7} fibroblasts were transduced with the Brunello library (four sgRNAs per gene), genomic DNA was collected at days 0 and 15, and gRNA changes were measured (left). Right, enrichments. Two technical replicates. Data represent the log_2-transformed fold change in read

counts before and after enrichment (PinAPL-Py). gRNAs with a $\log_2$-transformed fold change of >2 are shown (Supplementary Information). Two independent experiments were performed. **b**, RNA-seq analysis of ISGs of growing (PD61), crisis (PD105) and pre-crisis (PD100) $IMR90^{E6E7}$ fibroblasts that were transfected with two individual short interfering RNAs (siRNAs) against *ZBP1* (si*ZBP1*) or control (siCtrl). Significantly upregulated (crisis versus growing) and downregulated ISGs (si*ZBP1* versus siCtrl) with a fold change of >1.5 and a false-discovery rate (FDR)-adjusted $P < 0.05$ are shown (the complete list of ISGs used is from ref. [51]). One experiment was performed. **c**, Representative IncuCyte images of pre-crisis (PD94) *ZBP1*-knockout $IMR90^{E6E7}$ fibroblasts reconstituted with either WT ZBP1(S)–Flag or ZBP1(S)–Flag containing point mutations in Zα2, after 12 h of incubation (left). Dead cells are labelled with Cytotox green and nuclei are labelled with NucLight red. Scale bar, 150 μm. Three independent experiments were performed. Right, IncuCyte analysis of cell death of pre-crisis (PD94) *ZBP1*-knockout $IMR90^{E6E7}$ fibroblasts reconstituted with either WT ZBP1(S)–Flag or ZBP1(S)–Flag containing point mutations in the Zα2, RHIM1 or RHIM2 domains, measured in real time by Cytotox green. Data are mean ± s.e.m. from technical replicates. Statistical analysis was performed using one-way analysis of variance (ANOVA); $^{***}P < 0.001$. Three independent experiments were performed. **d**, RT–qPCR analysis of ISGs. Expression levels were normalized to control cells with the empty vector. Data are mean ± s.d. from technical replicates. *n* values indicate the number of technical replicates. Statistical analysis was performed using one-way ANOVA; NS, not significant; $^{***}P < 0.001$. Three independent experiments were performed.

Source data

ZBP1 can bind to left-handed Z-nucleic acid in a structure-specific manner[14,15,16,17]. Originally described as DNA sensor[18], ZBP1 is now appreciated to also sense RNAs that might exist in the Z conformation[14,19,20,21,22,23,24,25,26]. Alternative splicing of human *ZBP1* generates two main isoforms: full-length ZBP1 (ZBP1(L)) contains two N-terminal nucleic-acid-binding domains (Zα1 and Zα2) followed by RHIM domains (RHIM1 and RHIM2) and a structurally undefined C terminus; the short isoform (ZBP1(S)) lacks Zα1 but retains Zα2[27,28] (Extended Data Fig.

4a). *ZBP1* mRNA levels were induced in crisis, consistent with enhanced IFN signalling (Extended Data Fig. 4b). ZBP1(S) protein was upregulated, whereas ZBP1(L) was less abundant (Extended Data Fig. 4c,d). Increased ZBP1 expression correlated with the activation of type I IFN signalling, marked by TBK1, IRF3 and STAT1 phosphorylation (Extended Data Fig. 4c). LC3 lipidation and accumulation of LC3-II indicated ongoing autophagy[1] (Extended Data Fig. 4c). Targeting the major innate immune sensors and adapters revealed that ZBP1 induction was dependent on CGAS, STING and IFN, suggesting that CGAS–STING could drive an initial transcriptional induction of ZBP1 and prime cells to respond to aberrant accumulation of additional cytosolic immunostimulatory nucleic acid species (Extended Data Fig. 4e).

ZBP1 signalling requires Zα2 and RHIM1 domains

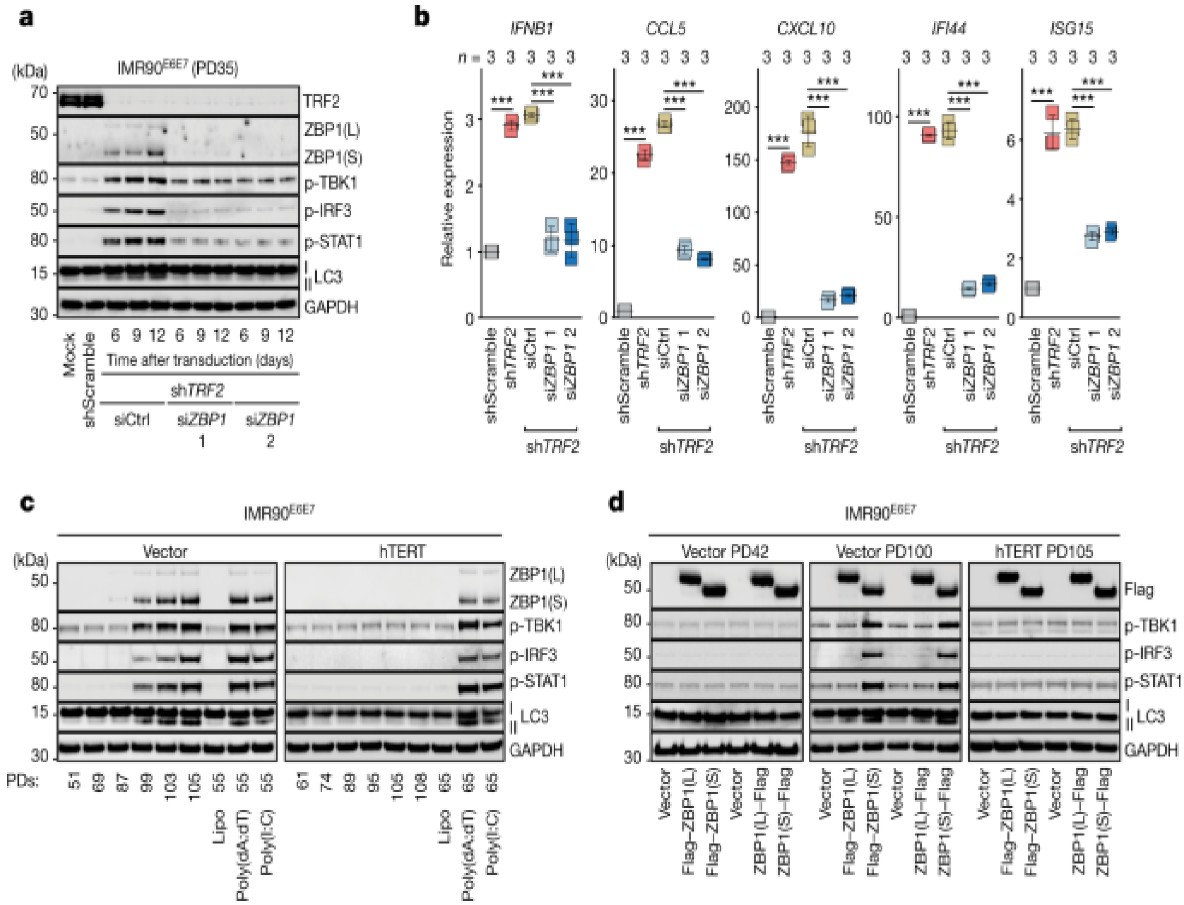
To address which domains of ZBP1 are critical for signalling during crisis, we disrupted the nucleic-acid-binding and RHIM functions by point mutations or truncations (Extended Data Fig. 5a). *ZBP1*-knockout cells were reconstituted with wild-type (WT) or mutant forms of ZBP1(S) (Extended Data Fig. 5b,c) and tested for their susceptibility to induce ISGs and undergo cell death. Pre-crisis cells expressing WT ZBP1(S) showed increased expression of ISGs concomitant with elevated cell death (Fig. 1c,d). However, cells reconstituted with ZBP1(S) containing mutations in key conserved residues involved in nucleic acid binding (N141A, Y145A)[29] or RHIM1-based interaction (I206A, Q207A, I208A, G209A)[30,31] did not induce ISGs and cell death (Fig. 1c,d). Mutating RHIM2 (V264A, Q265A, L266A, G267A)[30,31] did not prevent activation of ISGs or cell death (Fig. 1c,d). Comparable patterns were observed with ZBP1(S) truncation mutants (Extended Data Fig. 5d). Examining pathways downstream of ZBP1(S) revealed no role for RIPK1- or RIPK3-dependent PANoptosis, whereas autophagy was induced under these circumstances[1] (Extended Data Fig. 6a–d). Accordingly, suppression of RIPK3/NLRP3/PYCARD/CASP1-mediated pyropotosis, RIPK1/FADD/CASP8-mediated apoptosis or RIPK3/MLKL-mediated necroptosis did not protect cells from ZBP1(S)-induced cell death (Extended Data Fig. 6e). By contrast, elimination of components required

for IFN signalling or the autophagy machinery phenocopied the protective effects of ZBP1 depletion (Extended Data Fig. 6e). These findings revealed a mechanism by which ZBP1(S) drives cell death in a manner that is dependent on autophagy and type I IFN activities.

Innate signalling requires telomere dysfunction

Replicative crisis is a telomere-dependent program. We therefore depleted the telomere protection factor TRF2 in growing IMR90^{E6E7} cells, which resulted in telomere deprotection and fusion (Extended Data Fig. 6f–h), accompanied by ZBP1(S) upregulation and a type I IFN response through TBK1 phosphorylation (Fig. 2a). Depletion of ZBP1 dampened TBK1–IRF3 signalling, reduced ISG expression and attenuated autophagy, without affecting telomere fusions (Fig. 2a,b and Extended Data Fig. 6i,j). Expression of the catalytic subunit of telomerase (hTERT) in growing cells resulted in telomere maintenance and continued growth past crisis (Extended Data Fig. 7a,b). These immortalized cells did not express ZBP1 or activate IFN signalling (Fig. 2c and Extended Data Fig. 7c), whereas control cells entered crisis, upregulated ZBP1(S) and activated the TBK1–IRF3–IFN signalling axis (Fig. 2c and Extended Data Fig. 7c). hTERT expression did not affect the ability of cells to launch an IFN response or stimulate autophagy when treated with exogenous double-stranded RNA or DNA (Fig. 2c). Expression of either ZBP1(L) or ZBP1(S) in growing or telomerase-positive cells did not activate the IFN pathway (Fig. 2d (left and right)). ZBP1(S) potentiated an IFN response only when expressed in pre-crisis cells with short telomeres (Fig. 2d (middle)), suggesting the requirement for an additional immunostimulatory molecule specific for telomere dysfunction. In conclusion, two stimuli are required for launching a ZBP1-dependent IFN response in crisis: (1) upregulation of ZBP1(S) by CGAS–STING, and (2) a signal provided by dysfunctional telomeres.

Fig. 2: Telomere dysfunction is required for ZBP1-dependent IFN signalling.



a, Immunoblot analysis of growing (PD35) IMR90^{E6E7} fibroblasts expressing non-targeting control short hairpin RNA (shRNA) or shRNA against *TRF2* (sh*TRF2*). sh*TRF2* cells were transfected with either two individual siRNAs against *ZBP1* (si*ZBP1*) or non-targeting control siRNA at day 4 after shRNA transduction. Protein extracts were collected at days 6, 9 and 12 as shown in the timeline of the experiment in Extended Data Fig. 6i. Mock represents non-transduced cells. GAPDH was the loading control. Two independent experiments were performed. p, phosphorylated. **b**, RT–qPCR analysis of ISGs in growing (PD35) IMR90^{E6E7} fibroblasts expressing non-targeting control shRNA or shRNA against *TRF2*. sh*TRF2* cells were transfected with either two individual siRNAs against *ZBP1* or non-targeting control siRNA (siCtrl) at day 4 after shRNA transduction. RNA extracts were collected at day 12. Expression levels were normalized to control cells with non-targeting shRNA. Data are mean ± s.d. of technical replicates. *n* values indicate the number of technical replicates. Statistical analysis was performed using one-way ANOVA; ***$P < 0.001$. Three independent experiments were performed. **c**, Immunoblot analysis of IMR90^{E6E7}

fibroblasts expressing either empty vector or hTERT at the indicated PDs. Growth curves and telomere restriction fragment (TRF) analysis are shown in Extended Data Fig. 7a,b. Cells stimulated with 2 μg ml^{-1} of poly(deoxyadenylic-deoxythymidylic) (poly(dA:dT)) or poly(inosinic:cytidylic) (poly(I:C)) for 24 h were used as positive controls. GAPDH was used as the loading control. Two independent experiments were performed. **d**, Immunoblot analysis of growing (PD42), pre-crisis (PD100) and telomerase-positive (PD105) $IMR90^{E6E7}$ fibroblasts expressing ZBP1(L) or ZBP1(S). Flag tag was added to either the N terminus or the C terminus. GAPDH was the loading control. Two independent experiments were performed. Lipo, lipofectamine.

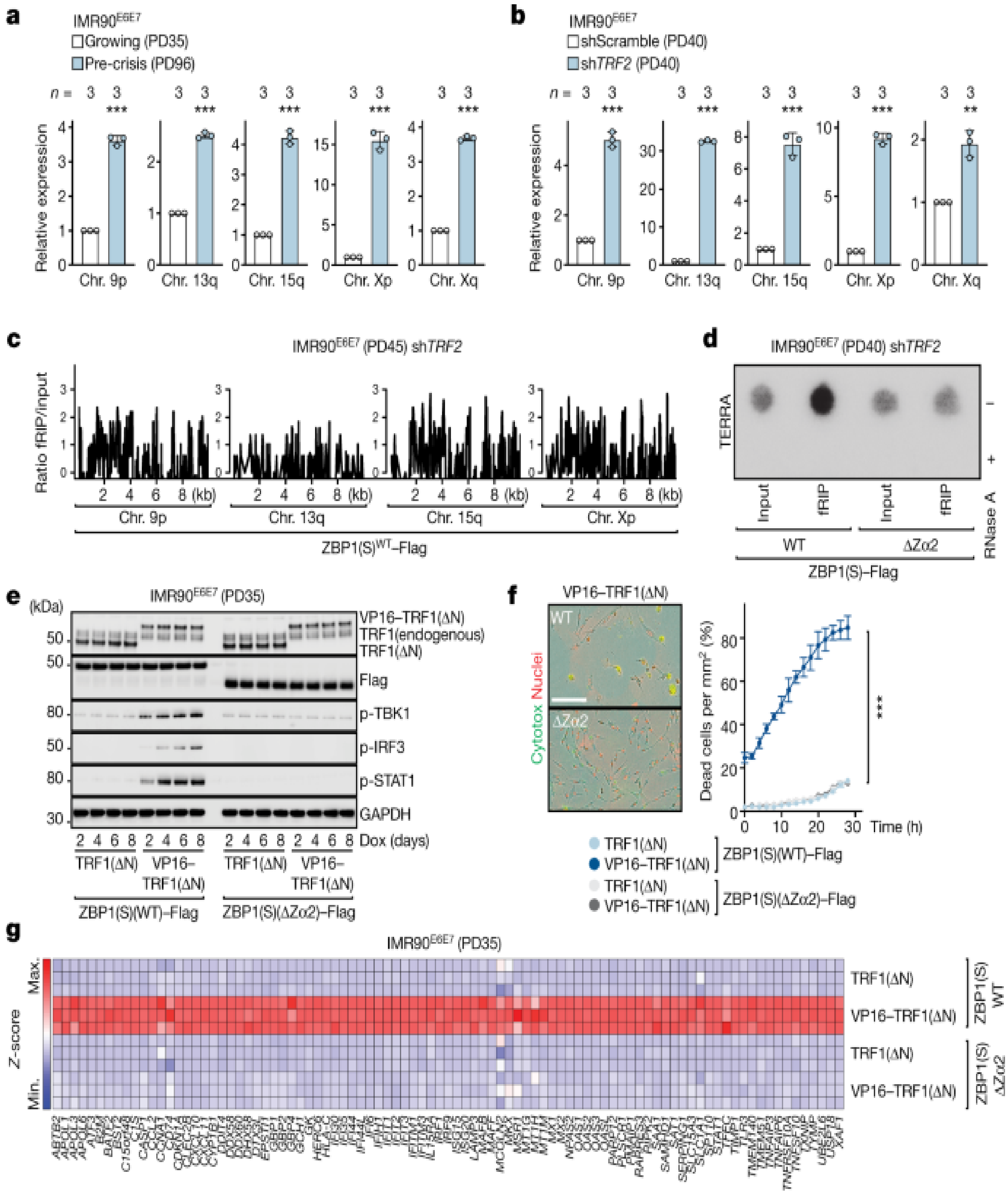
Source data

ZBP1 detects TERRA-derived RNA

The telomere-driven IFN response requires the nucleic-acid-binding function of ZBP1 (Fig. 1d and Extended Data Fig. 5d). Furthermore, ZBP1 expression in crisis cells was dependent not only on the CGAS–STING pathway, but also on mitochondrial antiviral signalling protein (MAVS) (Extended Data Fig. 4e)—an innate immune adapter that is located at the mitochondrial outer membrane (MOM). MAVS can be activated by the viral RNA sensors retinoic-acid-inducible gene I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5)[32,33,34,35], which did not have a role in ZBP1 upregulation during crisis (Extended Data Fig. 4e). Dysfunctional telomeres are actively transcribed into long non-coding RNA species termed TERRA, resulting in subtelomeric RNA sequences followed by a variable number of telomeric UUAGGG repeats[36,37]. RNA-dot blot and quantitative PCR with reverse transcription (RT–qPCR) analysis showed enhanced TERRA transcription from critically short telomeres in crisis or telomeres depleted of TRF2[38] (Fig. 3a,b and Extended Data Fig. 8a,b). Given that long non-coding RNAs interact with cytosolic sensors to regulate immune responses[39,40] and that TERRA has been linked to inflammation previously[41,42], we reasoned that TERRA could be the immunostimulatory nucleic acid species that is recognized by ZBP1(S). We first depleted TRF2 in growing $IMR90^{E6E7}$ fibroblasts expressing WT ZBP1(S) and then

performed formaldehyde cross-linking combined with ZBP1-associated RNA immunoprecipitation (Fig. 3c and Extended Data Fig. 8c,d). Formaldehyde cross-linking combined with RNA immunoprecipitation–sequencing (fRIP–seq) reads were mapped, and enrichment profiles were generated by calculating the fold change of the read counts between the fRIP and input samples. The analysis revealed reads for subtelomeres, with sharp peaks appearing at the telomere-proximal region located 5–10 kb upstream of TTAGGG repeats, implying that ZBP1 senses TERRA-containing transcripts (Fig. 3c and Extended Data Fig. 8c,d). To confirm ZBP1–TERRA interactions, we suppressed TRF2, immunoprecipitated WT or mutant ZBP1 lacking Zα2 (ZBP1(ΔZα2)) and assessed the presence of TERRA using an RNA-dot blot analysis (Fig. 3d and Extended Data Fig. 8e). A strong signal for TERRA was detected in immunoprecipitates from WT ZBP1 cells but not in those expressing mutant ZBP1 (Fig. 3d). Treatment with RNase A confirmed RNA-specific signals (Fig. 3d). These results indicate that telomere dysfunction contributes to the accumulation of TERRA molecules that physically interact with ZBP1 through its Zα2 domain.

Fig. 3: ZBP1 senses TERRA.



a,**b**, RT–qPCR analysis of TERRA transcripts. RNA from growing and pre-crisis IMR90^{E6E7} fibroblasts (**a**) and growing IMR90^{E6E7} fibroblasts expressing control shRNA or shRNA against *TRF2* (**b**). RNA was collected at day 4 after transduction, normalized to growing (left) or shScramble (right). Data are mean ± s.d. from technical replicates. *n* values indicate the number of technical replicates. Statistical analysis was performed using two-

tailed Student's *t*-test. ***$P < 0.001$. Three independent experiments. **c**, fRIP–seq enrichment of ZBP1-associated RNA transcripts within 10 kb upstream of the subtelomere–telomere boundary. fRIP analysis was performed on growing (PD45) IMR90^{E6E7} fibroblasts expressing WT ZBP1(S)–Flag and transduced with either control shRNA or sh*TRF2*. Immunoprecipitation at day 12 after shRNA transduction. Validation is shown in Extended Data Fig. 8c. fRIP and input samples were normalized to the same sequencing depth and average ratios of individual subtelomeres shown in Extended Data Fig. 8d. Data are log_2-transformed. Two independent experiments were performed. **d**, RNA-dot blot from anti-Flag immunoprecipitates (fRIP) or total lysates (input) using ^{32}P-dCTP-labelled probes targeting TERRA transcripts. fRIP analysis of growing (PD40) IMR90^{E6E7} fibroblasts expressing either WT ZBP1(S)–Flag or mutant ZBP1(S)–Flag lacking Zα2 with sh*TRF2*. Immunoprecipitation was performed at day 12 after shRNA transduction. Validation is shown in Extended Data Fig. 8e. RNase A treatment was used as a contamination control. Two independent experiments were performed. **e**, Immunoblot analysis of growing (PD35) IMR90^{E6E7} fibroblasts (WT ZBP1(S)–Flag or ZBP1(S)–Flag lacking Zα2) with either TRF1(ΔN) or VP16–TRF1(ΔN). Extracts were collected as indicated after doxycycline (Dox) (1 μg ml^{-1}) treatment. Two independent experiments were performed. **f**, IncuCyte images at day 50 (the growth curve is shown in Extended Data Fig. 10a) after 24 h of incubation (left). Cells are labelled as described in Fig. 1c. Scale bar, 200 μm. Three independent experiments were performed. Right, cell death by Cytotox green incorporation. Data are mean ± s.e.m. from technical replicates. Statistical analysis was performed using one-way ANOVA. ***$P < 0.001$. Three independent experiments were performed. **g**, ISG RNA heat maps. RNA was collected at day 10 after doxycycline treatment. Significantly upregulated ISGs (fold change > 3; FDR-adjusted $P < 0.05$) are shown. One experiment was performed. Chr., chromosome.

Source data

To evaluate the role of TERRA–ZBP1 interactions in innate immune activation, we expressed a mutant form of TRF1 lacking the N terminus (TRF1(ΔN)) alone or fused to the VP16 transcriptional activation domain (VP16–TRF1(ΔN))[42] in growing IMR90^{E6E7} fibroblasts. Stable expression

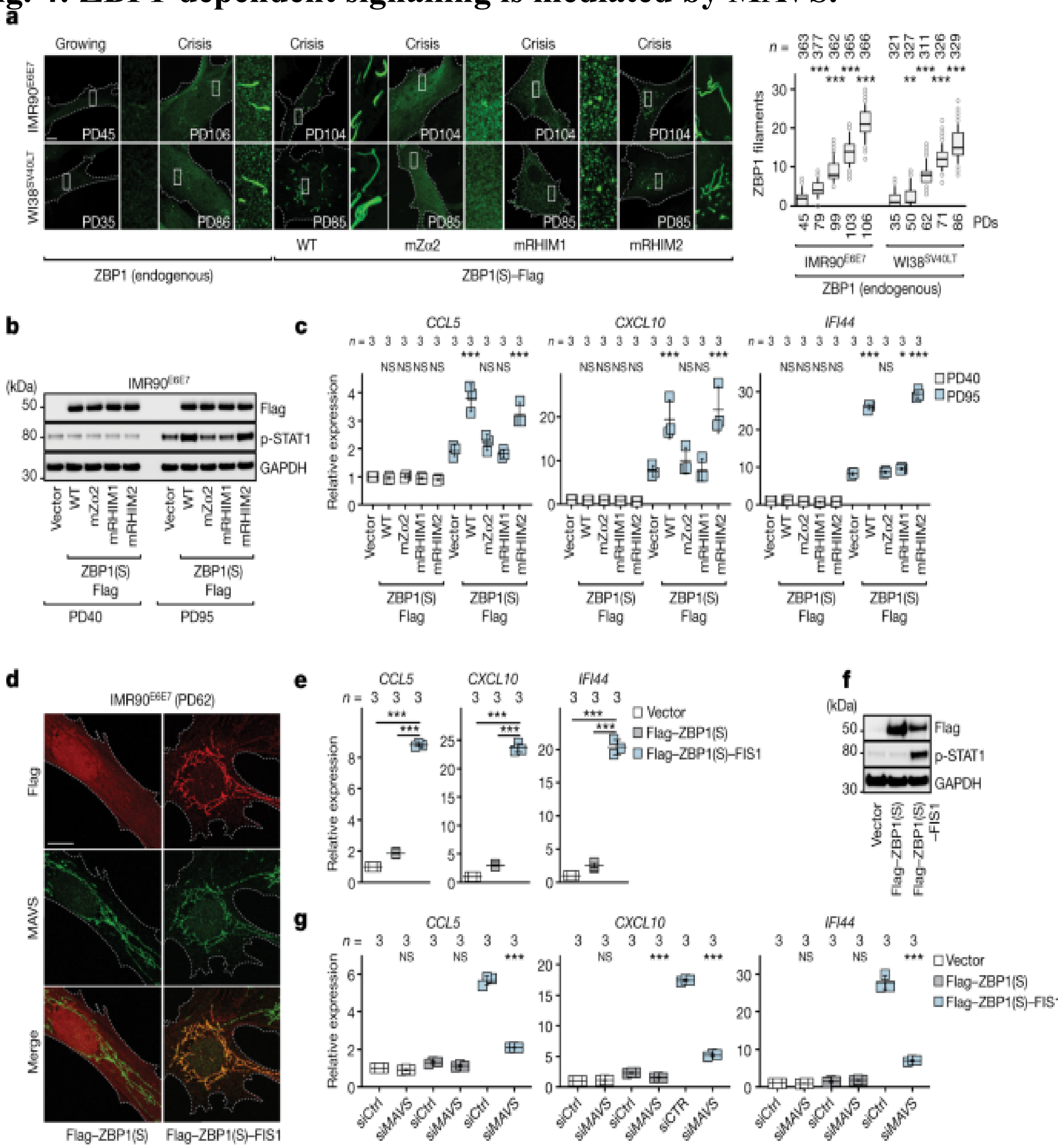
of these proteins was confirmed, and upregulation of TERRA was validated (Fig. 3e and Extended Data Fig. 9a–c). WT ZBP1-expressing cells responded to TERRA induction by activating the IFN signalling pathway and ISG expression (Fig. 3e,g). Accordingly, cells entered a crisis-like state associated with cell death and elevated ISG levels (Fig. 3f and Extended Data Fig. 10a–d). No loss of telomere protection, no activation of ATM- and ATR-dependent DNA-damage response and no formation of fused chromosomes were observed (Extended Data Fig. 11a,b). Partial depletion of TERRA in cells expressing WT ZBP1 delayed STAT1 activation (Extended Data Fig. 10e–g). Cells expressing mutant ZBP1 with disrupted nucleic-acid-binding activity were insensitive to TERRA and did not activate the innate immune response (Fig. 3e–g and Extended Data Fig. 10a–d). These findings demonstrate a mechanism of innate immune signalling triggered by dysfunctional telomeres, in which TERRA functions as the messenger molecule that binds to ZBP1(S) in the cytoplasm to promote an enhanced type I IFN response.

ZBP1 forms filaments at mitochondrial membranes

To decipher the mechanism of ZBP1 signalling, we performed immunostaining of ZBP1 in crisis cells, which revealed filamentous structures in the cytoplasm (Fig. 4a). ZBP1 lacking functional Zα2 and RHIM1 domains did not form these filaments, whereas no significant effect was observed after disruption of RHIM2 (Fig. 4a). A lack of filaments correlated with impaired type I IFN signalling (Fig. 4b,c), indicating that the assembly of ZBP1 filaments is essential for innate immune signalling and requires both TERRA binding through Zα2 and self-oligomerization through homotypic RHIM1 interactions. Multimerization is used by several sensors for signal propagation, including RIG-I[43] and MDA5[44], which translocate to mitochondria to interact and activate MAVS[32,33,34,35]; however, they often form aggregated structures rather than filaments[45], suggesting an alternative oligomerization mechanism for ZBP1. We found co-localization of ZBP1 filaments with a subset of mitochondria based on MOM (TOM20) and matrix (MitoTracker and TFAM) markers (Extended Data Fig. 11c). Inhibition of mtDNA replication with 2′,3′-dideoxycytidine[46] had no effect

on the number of ZBP1 filaments even though mtDNA and mtRNA levels were significantly reduced (Extended Data Fig. 11d). Depletion of either SUV3 or PNPase enzymes, components of the mitochondrial degradosome[47], led to significant accumulation of mtRNA without altering ZBP1 filament formation (Extended Data Fig. 11e). We therefore concluded that ZBP1-mediated innate immune response requires formation of filaments on mitochondria, independently of sensing mtDNA or mtRNA.

Fig. 4: ZBP1-dependent signalling is mediated by MAVS.



a, Growing and crisis IMR90^{E6E7} and WI38^{SV40LT} fibroblasts, and crisis IMR90^{E6E7} and WI38^{SV40LT} fibroblasts expressing WT ZBP1(S)–Flag or ZBP1(S)–Flag containing mutations in Zα2, RHIM1 or RHIM2, immunostained against endogenous ZBP1 or Flag (left). Scale bar, 10 μm. Three independent experiments were performed. Right, box and whisker plots of endogenous ZBP1 filaments. The median (centre line), first and third quartiles (box limits) and 10th and 90th percentiles (whiskers) are shown. *n* values indicate the cell number. Statistical analysis was performed using one-way ANOVA; $^{**}P < 0.01$; $^{***}P < 0.001$. Three independent experiments were performed. The white boxes indicate higher-magnification images. **b**, Immunoblot analysis of growing and pre-crisis IMR90^{E6E7} fibroblasts expressing empty vector, WT ZBP1(S)–Flag or ZBP1(S)–Flag containing domain point mutations. GAPDH was used as the loading control. Two independent experiments were performed. **c**, RT–qPCR analysis of ISGs in growing and pre-crisis IMR90^{E6E7} fibroblasts. Expression levels were normalized to growing cells expressing the empty vector. Data are mean ± s.d. of technical replicates. *n* values indicate the number of technical replicates. Statistical analysis was performed using one-way ANOVA. NS, not significant; $^{***}P < 0.001$. Three independent experiments were performed. **d**, Growing IMR90^{E6E7} fibroblasts expressing Flag–ZBP1(S) or Flag–ZBP1(S)–FIS1 immunostained for Flag and MAVS. Scale bar, 10 μm. Three independent experiments were performed. **e**, RT–qPCR analysis of ISGs. RNA extracts were collected 10 days after transduction. Expression levels were normalized to the empty vector. Data are mean ± s.d. from technical replicates. *n* values indicate the number of technical replicates. Statistical analysis was performed using one-way ANOVA. $^{***}P < 0.001$. Three independent experiments were performed. **f**, Immunoblot analysis of growing IMR90^{E6E7} fibroblasts as described in **e**. GAPDH was used as the loading control. Two independent experiments were performed. **g**, RT–qPCR analysis of ISGs in growing (PD66) IMR90^{E6E7} fibroblasts. Cells were transfected with si*MAVS* or non-targeting control siRNA at day 10 after transduction. RNA extracts were collected at day 14. Expression levels were normalized to empty vector siCtrl cells. Data are mean ± s.d. from technical replicates. *n* values indicate the number of technical replicates. Statistical analysis was performed using one-way ANOVA. NS, not significant; $^{***}P < 0.001$. Three independent experiments were performed.

Source data

MAVS is required for a ZBP1-mediated IFN response

Given that MAVS, a crucial adapter for RNA sensors anchored to the MOM, was enriched in our CRISPR–Cas9 knockout screen (Supplementary Information), and its depletion prevented IFN-dependent ZBP1 induction during crisis (Extended Data Fig. 4e), we reasoned that ZBP1-driven cell death could be mediated through MAVS. Indeed, depletion of MAVS prevented ISG expression in response to ZBP1 and TERRA co-expression independently of RIG-I and MDA5 (Extended Data Fig. 12a–c). Deletion of *MAVS* also reverted ISG expression during crisis, prevented autophagy upregulation and reduced the frequency of cell death (Extended Data Fig. 12d–f), suggesting that ZBP1 filaments at the mitochondria could be the critical event downstream of telomere stress signalling that leads to MAVS activation. To test this, ZBP1(S) was fused to the mitochondrial-targeting sequence of FIS1 to direct it specifically to the MOM in growing cells with functional telomeres (Extended Data Fig. 13a). ZBP1(S)–FIS1 colocalized with MAVS, triggered an IFN response, stimulated autophagy and induced cell death in fibroblasts and epithelial cells (Fig. 4d–f and Extended Data Fig. 13c,d). ISG induction by mitochondria-targeted ZBP1(S)–FIS1 was dependent on MAVS (Fig. 4g and Extended Data Fig. 13b). Finally, as shown by suppression of IFNAR2, cell death by autophagy required a secondary signalling pathway downstream of ZBP1–MAVS, by which secreted IFNs bind to and activate their cognate IFN receptor complexes in an autocrine and paracrine manner (Extended Data Fig. 13c,d). Together, these results suggest that ZBP1 signalling in crisis cells involves binding to immunostimulatory TERRA, followed by conformational changes allowing cytosol-to-MOM translocation and oligomerization to activate MAVS. We propose that these properties of ZBP1 serve as a signal-amplification mechanism enabling TERRA molecules to trigger a second wave of IFN signalling that, together with autophagy activation, causes cell death during crisis.

Although ZBP1 functions as an RNA sensor with the ability to activate a MAVS and IFN signalling response, it differs from RIG-I and MDA5 in

several aspects. First, ZBP1 filament formation occurs at the mitochondrial surface, whereas RIG-I[43,48] and MDA5[44,49,50] undergo oligomerization along double-stranded RNA structures. Second, ZBP1 lacks the caspase recruitment domain (CARD), whereas RIG-I and MDA5 activate MAVS through CARD–CARD interactions[35]. Finally, we propose that TERRA transcripts constitute a structure-specific ligand for ZBP1, with low or no binding affinity to RIG-I and MDA5. These differences may prime ZBP1 to propagate the inflammatory signalling cascade specifically in response to telomere dysfunction, revealing a unique mechanism of MAVS-dependent IFN activation and subsequent cell death by autophagy.

Our study reveals a mechanism for telomere-mediated tumour suppression, whereby dysfunctional telomeres in crisis stimulate two intertwined cytosolic nucleic-acid-sensing pathways and trigger a lethal IFN response (Extended Data Fig. 13e). We propose that the breakage of fused telomeres and the subsequent release of nuclear DNA into the cytoplasm drives initial activation of the CGAS–STING pathway and expression of ISGs, including *ZBP1*. However, crisis-associated cell death requires additional downstream activation of ZBP1(S) by TERRA and formation of ZBP1 filaments on mitochondria. This promotes an inflammatory loop leading to the expression of an ISG profile that, in concert with activation of autophagy through an as yet undetermined pathway, drives cell death specifically in replicative crisis. The simultaneous activation of CGAS–DNA-sensing and ZBP1–RNA-sensing pathways enables the innate immune system to orchestrate an efficient, type-I-IFN-dependent cell death response to eliminate precancerous cells with unstable telomeres. These findings highlight a synergy between critically short telomeres, mitochondria and innate immunity that has evolved to prevent age-associated cancer initiation in humans.

Methods

Cell culture

IMR90 (CCL-186) and WI38 (AG06814-N) fibroblasts were purchased from ATCC and the Coriell Institute for Medical Research, respectively. IMR90 and WI38 fibroblasts were grown under 7.5% CO_2 and 3% O_2 in

GlutaMax-DMEM (Gibco, 10569-010) supplemented with 0.1 mM non-essential amino acids (Corning, 25-025-Cl) and 15% fetal bovine serum (VWR/Seradigm, 97068-085, 311K18). Human mammary epithelial cells (HMECs; CC-2551) were purchased from Lonza and were grown under 5% CO_2 and 3% O_2 using the mammary epithelial cell medium complete kit (Lifeline Cell Technology, LL-0061). The number of PDs was calculated at each passage using the following equation: PD = log[number collected/seeded number]/$\log_2$. Cells have been tested to be free of mycoplasma.

Plasmids

Non-coding scramble shRNA was obtained from D. Sabatini through Addgene (plasmid 1864)[52]. *TRF2* shRNA in pLKO.1 was obtained from Open Biosystems. HPV16 E6E7, SV40 LT and H2B-mCherry in pLXSN3 and hTERT in pBabe were obtained from Karlseder laboratory stock[2]. To generate shZBP1 1, oligos (forward, 5′-CCGGCCAAGTCCTCTACCGAATGAACTCGAGTTCATTCGGTAGAGGACTTGGTTTTTG-3′; and reverse, 3′-AATTCAAAAACCAAGTCCTCTACCGAATGAACTCGAGTTCATTCGGTAGAGGACTTGG-5′) were annealed by temperature ramp from 100 °C to 25 °C and cloned in a AgeI- and EcoRI-digested non-coding scramble shRNA vector using T4 DNA ligase. *ZBP1* shRNA in pLKO.1 (shZBP1 2) was obtained from Sigma-Aldrich (TRCN0000123050).

To generate pooled knockout cell populations and knockout clones lentiCRISPRv2 was used. The plasmid was obtained from F. Zhang through Addgene (plasmid 52961)[53] and the guide target sequence was cloned as phosphorylated adapter in the Esp3I-digested lentiCRISPRv2 vector. Individual guide RNAs (gRNAs) for the targeting of corresponding genomic exons were designed using the CRISPOR web tool (http://crispor.org). The guide RNA sequences are as follows: sgZBP1 1, 5′-CACCTGGTGCCATTGAAGGG-(PAM)-3′; sgZBP1 2, 5′-GGACGATTTACCGCCCAGGT-(PAM)-3′; sgZBP1 3, 5′-TGGGACACAGCAATGAGATG-(PAM)-3′; sgGFP, 5′-GAAGTTCGAGGGCGACACCC-(PAM)-3′; sgLUC (sgrenilla_luciferase), 5′-GGTATAATACACCGCGCTAC-(PAM)-3′; sgMAVS, 5′-

AAGTTACCCCATGCCTGTCC-(PAM)-3′; and sgSTING, 5′-AATATGACCATGCCAGCCCA-(PAM)-3′. *ZBP1*-knockout clones were generated using sgZBP1 2 5′-GGACGATTTACCGCCCAGGT-(PAM)-3′. After transduction, cells were selected with 1 μg ml^{-1} puromycin for 1 week and individual clones were isolated using the array dilution method in 96 wells. Clones were tested by western blotting for the absence of ZBP1 protein staining.

All *ZBP1* expression plasmids used here are derivates of pLenti-CMV-GFP-Hygro-(656-4) obtained from E. Campeau and P. Kaufman through Addgene (plasmid 17446)[54]. The sequence of full-length human *ZBP1* transcript variant 1 (GenBank: NM_030776.3) was obtained as gblock from IDT. The empty vector control pLenti-MCS-hygro was cloned in three steps: first, the pLenti-CMV-GFP-Hygro-(656-4) vector was digested with XbaI and SalI and the GFP sequence was replaced with a multiple cloning site (MCS) adapter (MCS1_fw, 5′-CTAGACCTCAGGATCCCGGGACGCGTG-3′; and MCS2_rev, 5′-TCGACACGCGTCCCGGGATCCTGAGGT-3′) to generate pLenti-CMV-MCS-Hygro. Second, the CMV promoter in the pLenti-CMV-MCS-Hygro vector was excised using ClaI and BamHI and replaced with a second adapter (MCS2_fw, 5′-CGATCCTGAGGTCTAGAGAATTCG-3′; and MCS2_rev, 5′-GATCCGAATTCTCTAGACCTCAGGAT-3′). Finally, pLenti-MCS-WPRE-Hygro was digested with SalI and XhoI to remove the WPRE sequence followed by religation to generate pLenti-MCS-Hygro. To clone N-terminally 3×Flag-tagged *ZBP1* expression constructs, pLenti-CMV-3×Flag-$(GS)_5$-MCS-Hygro was generated first. For this, the Kozak sequence with a 3×Flag tag and a $(GS)_5$ linker was amplified from pLenti-FNLS-P2A-GFP-PGK-Puro (obtained from L. Dow through Addgene (plasmid 110869)[55] using primers 5′-CTAGACCTCAGGATCGCCACCATGGACTATAAGGAC-3′ and 5′-ACGCGTCCCGGGATCCAGAGCCGGACCCGCTCCCGGAGCCCTTATCGTCATCGTCTTTGTAATCAATATCATGAT-3′. The fragment was cloned into BamHI-linearized pLenti-CMV-MCS-Hygro using InFusion cloning (Takara Bio, 639650).

ZBP1 full-length (ZBP1(L)) and short isoform (ZBP1(S), also known as Delta-exon-2, Delta-Z-alpha or isoform 7) were amplified from full-length *ZBP1* gblock DNA using either the primer 5′-

GTCCGGCTCTGGATCCATGGCCCAGGCTCCTGCT-3′ or primer 5′-GTCCGGCTCTGGATCCATGGCCCAGGCTCCTGCTGACCCGGGCAGAGAAGCCGAGAGGCCCCAGC-3′ in combination with the primer 5′-ACGCGTCCCGGGATCCCTAAATCCCACCTCCCCACC-3′. *ZBP1* fragments were cloned into BamHI-linearized pLenti-CMV-3×Flag-$(GS)_5$-MCS-Hygro using InFusion cloning. To clone C-terminally 3×Flag-tagged *ZBP1* expression constructs, pLenti-CMV-MCS-$(GS)_5$-3×Flag-Hygro was generated first. For this, a $(GS)_5$ linker with a 3×Flag tag and STOP codon was amplified from pLenti-FNLS-P2A-GFP-PGK-Puro using the primers 5′-CGACTCTAGAGGATCCGGCTCCGGGAGCGGGTCCGGCTCTGACTATAAGGACCACGACGG-3′ and 5′-GAGGTTGATTGTCGACTCACTTATCGTCATCGTCTTTGTAATC-3′. The fragment was cloned into BamHI- and SalI-digested pLenti-CMV-GFP-Hygro-(656-4) using InFusion cloning.

ZBP1(L)-, ZBP1(S)- and ZBP1(S)-deletion mutants were PCR amplified from full-length *ZBP1* gblock DNA and the pLenti-CMV-3×Flag-$(GS)_5$-ZBP1(S)-Hygro plasmid, respectively, with the primers listed below. Fragments were cloned into BamHI-digested pLenti-CMV-MCS-$(GS)_5$-3×Flag-Hygro using InFusion cloning.

The primers for the C-terminal tagged *ZBP1* construct cloning were as follows: ZBP1(L) (amino acids 1–429) and ZBP1(S) (amino acids 1–11, 87–429) ZBP1, 5′-CGACTCTAGAGGATCCATGGCCCAGGCTCCTG-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′; ZBP1(S)-ΔZα2 (amino acids 167–429), 5′-CGACTCTAGAGGATCCATGCCAGAAGATTCTGGAAGAAGAGCAA-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′; ZBP1(S)-ΔZα2-ΔRHIM1-ΔRHIM2 (amino acids 278–429), 5′-CGACTCTAGAGGATCCATGCCGTCCGAGGGCCCT-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′; ZBP1(S)-ΔC (amino acids 1–11, 87–319), 5′-CGACTCTAGAGGATCCATGGCCCAGGCTCCTG-3′ and 5′-TCCCGGAGCCGGATCCGGCGGCTTCCCCCT-3′; ZBP1(S)-ΔRHIM1 (amino acids 1–11, 87–194, 220–429) N-terminal fragment, 5′-CGACTCTAGAGGATCCATGGCCCAGGCTCCTG-3′ and 5′-TCCCTGGAGGGTCCATTCTGGCAGATCA-3′ and C-terminal fragment

5′-TGGACCCTCCAGGGAGGACGGT-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′; ZBP1(S)-ΔRHIM2 (amino acids 1–11, 87–252, 278–429) N-terminal fragment, 5′-CGACTCTAGAGGATCCATGGCCCAGGCTCCTG-3′ and 5′-TCGGACGGCTGGGGCCCCCAGG-3′ and C-terminal fragment, 5′-GCCCCAGCCGTCCGAGGGCCCTGC-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′; and ZBP1(S)-ΔRHIM1-ΔRHIM2 (amino acids 1–11, 87–194, 278–429) N-terminal fragment, 5′-CGACTCTAGAGGATCCATGGCCCAGGCTCCTG-3′ and 5′-TCGGACGGGGGTCCATTCTGGCAGATCA-3′ and C-terminal fragment, 5′-TGGACCCCCGTCCGAGGGCC-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′. Zα2, RHIM1 and RHIM2 point mutations were introduced in the pLenti-CMV-ZBP1(S)-$(GS)_5$-3×Flag-Hygro vector by site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, 210518). The primers used here for site-directed mutagenesis were as follows: ZBP1(S)-mZα2 (N141A,Y145A), 5′-GGACAGCAAAAGATGTGGCCCGAGACTTGGCCAGGATGAAGAGCAGGC-3′ and 5′-GCCTGCTCTTCATCCTGGCCAAGTCTCGGGCCACATCTTTTGCTGTCC-3′; ZBP1(S)-mRHIM1 (I206A,Q207A,I208A,G209A), 5′-TGGATTTCCATTGCAAACTCCGAAGCCGCCGCGGCTGCACACGGGAACATCATTACAAGACAGAC-3′ and 5′-GTCTGTCTTGTAATGATGTTCCCGTGTGCAGCCGCGGCGGCTTCGGAGTTTGCAATGGAAATCCA-3′; and ZBP1(S)-mRHIM2 (V264A,Q265A,L266A,G267A), 5′-GCAGTCCATACTGAGACGGGCGGCGGCGGCACACAGCAATGAGATGAGGC-3′ and 5′-GCCTCATCTCATTGCTGTGTGCCGCCGCCGCCCGTCTCAGTATGGACTGC-3′.

To express untagged VP16-TRF1ΔN and TRF1ΔN from a doxycycline-inducible tight TRE promoter, VP16-TRF1ΔN and TRF1ΔN was subcloned into pLPT-Empty[56]. For this, VP16-TRF1ΔN and TRF1ΔN were PCR amplified from pINDUCER20-VP16–TRF1ΔN(44–439) and pINDUCER20-TRF1ΔN(44–439), respectively (both plasmids were kind gifts from P. M. Lieberman[57]), using either primer 5′-

TGGAGAATTGGCTAGCATGGCCCCCCCGACCG-3′ or primer 5′-TGGAGAATTGGCTAGCATGCTTCTCGAGTGCCAGGTGC-3′ in combination with primer 5′-CCCCAACCCCGGATCCTCAGTCTTCGCTGTCTGAGGAAATCAG-3′. Fragments were cloned into the NheI and BamHI-digested pLPT-Empty vector using InFusion cloning.

To localize ZBP1(S) to the mitochondrial outer membrane, the outer mitochondrial membrane localization (OMM) signal of FIS1 was fused to the C terminus of 3×Flag-ZBP1(S). The construct was cloned in three steps: first, pLenti-CMV-MCS-3×Flag-OMM-hygro was generated by amplifying the OMM of FIS1 with the primers 5′-CGATAAGGCCATCGTGGGAGGC-3′ and 5′-GAGGTTGATTGTCGACTCAGGATTTGGACTTGGACACAG-3′ from MAVS-Mito (obtained from J. Kagan through Addgene (plasmid 44556)[58] and (GS)5-linker-3×Flag with the primers 5′-CGACTCTAGAGGATCCGGCTCCGGGAGCGGGTCCGGCTCTGACTATAAGGACCACGACGG-3′ and 5′-CACGATGGCCTTATCGTCATCGTCTTTGTAATC-3′ from pLenti-FNLS-P2A-GFP-PGK-Puro[5]. The two fragments were cloned into BamHI/SalI-digested pLenti-CMV-GFP-Hygro(656-4) (see methods of ref. [4]) using InFusion cloning. Next, pLenti-CMV-MCS-(GS)5-3Flag-OMM-hygro was linearized with BamHI and ZBP1(S) that was amplified with the primers 5′-CGACTCTAGAGGATCCATGGCCCAGGCTCCTG-3′ and 5′-TCCCGGAGCCGGATCCAATCCCACCTCCCCACC-3′ from pLenti-CMV-3×Flag-(GS)5-ZBP1(S)-hygro was cloned into the linearized vector using InFusion cloning. Finally, pLenti-CMV-3×Flag-(GS)5-ZBP1(S)-OMM-hygro was generated by InFusion cloning using amplified fragments of pLenti-CMV-ZBP1(S)-(GS)5-3Flag-OMM with the primers 5′-GCCATCGTGGGAGGCATG-3′ and 5′-GGATCCTCTAGAGTCGGTGTCT-3′ and 3×Flag-(GS)5-ZBP1(S) with the primers 5′-GACTCTAGAGGATCCGCCACCATGGACTATAAGGACCACGAC-3′ and 5′-GCCTCCCACGATGGCAATCCCACCTCCCCACC-3′ from pLenti-CMV-3×Flag-(GS)5-ZBP1(S)-hygro.

T4 ligase (M0202S), T4 PNK (M0201S) and all restriction enzymes used, except for Esp3I (Thermo Scientific ER0452), were obtained from NEB. All

primers were ordered from Eton Bioscience. PCR reactions for cloning were performed using KOD Hot Start DNA Polymerase (Novagen, 71086) according to the manufacturer's protocol.

Cell death assays

For real-time assessment of cell death, 2,000 cells per well per condition in triplicate were seeded in 96-well plates (CytoOne, CC7682-7596) the day before the experiment. The next day, the medium was changed to medium containing 250 nM IncuCyte Cytotox green dye (Sartorius, 4633) and IncuCyte Nuclight Rapid red dye (1:100 dilution) (Sartorius, 4717). Cells were seeded in 96-well plates the day before the experiment and the medium changed to medium containing 250 nM IncuCyte Cytotox green dye on the day of the experiment. Cells were imaged with at least two fields per well every 2 h using the ×10 objective on the IncuCyte S3 or IncuCyte Zoom live-cell analysis system (Sartorius). The total number of nuclei and the Cytotox-green-dye-positive nuclei were quantified using the IncuCyte analysis software. The percentage of dead cells was calculated by dividing the Cytotox green dye positive nuclei by the total number of nuclei multiplied by 100. In Extended Data Figs. 6e and 13d, cells were transfected with siRNA 48 h before cell death measurements. Then, 24 h after transfection, the medium was changed to medium containing 250 nM IncuCyte Cytotox green dye (Sartorius, 4633) or IncuCyte Cytotox red dye (Sartorius, 4632). Then, 48 h after transfection, nuclei were stained with Hoechst (1 μg ml^{-1}), images were taken using the Revolve fluorescence microscope and analysed with CellProfiler v.4.2.1 using a customized pipeline. Total nuclei and Cytotox-dye-positive nuclei were segmented using an integrated intensity-based object detection module. The percentage of dead cells was calculated by dividing the number of Cytotox-dye-positive nuclei by the total number of nuclei multiplied by 100.

Retroviral and lentiviral transduction

Lentiviral and retroviral particles were produced by the laboratory. Production of lentivirus was performed as described previously[56]. In brief, HEK293T (ATCC, CRL-11268) cells were transfected with 7 μg of DNA using Lenti-X Packaging Single-Shot system (Clontech, 631276). Then, 48 h

after transfection, viral supernatant was collected, supplemented with serum and used for transduction in the presence of Lenti-Blast (Oz Biosciences, LB00500). To produce retrovirus, Phoenix cells were transfected with 20 μg of DNA using 100 μM of chloroquine. Then, 5 h after transfection, fresh medium was added. The viral supernatant was collected 24 h later and used for transduction in the presence of polybrene 4 μg ml^{-1}. Then, 48 h after infection, cells were washed and selected with 1 μg ml^{-1} puromycin, 600 μg ml^{-1} G418, or 90 μg ml^{-1} hygromycin. $IMR90^{E6E7}$ and $WI38^{SV40LT}$ fibroblasts were subjected to long-term culturing under antibiotic selection.

Transfections

DNA transfections were performed using the Lipofectamine 3000 kit (Thermo Fisher Scientific, 1857482) according to the manufacturer's instructions. siRNA transfections were performed using the Lipofectamine RNAiMAX kit (Thermo Fisher Scientific, 13778030) according to the manufacturer's instructions.

Western blotting

Western blots were performed as described previously[59]. In brief, cells were lysed in NuPage LDS sample buffer (Invitrogen, NP0007) at 1×10^4 cells per μl. Proteins were resolved using NuPage Bis-Tris gel electrophoresis (Invitrogen, NP0343, NP0321 or WG1402) and transferred to nitrocellulose membranes (Amersham, 10600037). Secondary antibodies were peroxidase-conjugated anti-mouse IgG (Amersham, NXA931V) or anti-rabbit IgG (Amersham, NA934V). Peroxidase activity was detected using an ECL kit (Prometheus, 20-302B) and the Syngene G-Box imager. Primary antibodies are listed below.

Metaphase spreads for quantification of chromosome fusions

Metaphase spread preparation was performed as described[60]. In brief, cells were treated with 0.2 μg ml^{-1} of Colcemid (Gibco, 15212-012) for 3 h, collected and incubated with hypotonic solution (75 mM of KCl) for 7 min at 37 °C. The cell suspension was centrifuged, washed in fixative solution

(3:1 methanol:glacial acetic acid), dropped onto superfrost microscope slides (VWR, 48311-703) and air dried overnight. For telomere staining, cells were fixed in 4% formaldehyde in PBS for 10 min, dehydrated in a series of ethanol baths (70%, 90% and 100%) for 3 min each and air-dried for 20 min. Cells were then covered in 0.3 ng μl^{-1} of the telomeric (PNA Bio, F1004) PNA probes diluted in hybridization solution (70% deionized formamide, 0.25% blocking reagent (NEN), 10 mM Tris pH 7.5). The samples were heated for 5 min at 80 °C, incubated 2 h at room temperature, washed twice in 70% formamide and 10 mM Tris-HCl pH 7.5 and three times in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.08% Tween-20. Slides were finally mounted in ProLong Diamond with DAPI (Invitrogen, P36971).

Immunofluorescence-FISH analysis of interphase nuclei and metaphase spreads

For interphase nuclei, cells were seeded onto glass coverslips 24 h before the experiment as described previously[60]. Cells were washed in PBS, fixed in 4% formaldehyde in PBS for 10 min and permeabilized in 0.1% Triton X-100 in PBS for 10 min. For metaphase spreads, cells were treated with 20 ng ml^{-1} of Colcemid (Gibco, 15212-012) for 1 h, collected and incubated in hypotonic solution (27 mM KCl, 6.5 mM tri-sodium citrate) for 5 min. The cell suspension was cytocentrifuged, fixed in 4% formaldehyde in PBS for 10 min and permeabilized in KCM buffer (120 mM KCl, 20 mM NaCl, 10 mM Tris pH 7.5, 0.1% Triton X-100) for 10 min. In both settings, the samples were incubated in blocking buffer (20 mM Tris pH 7.5, 2% BSA, 0.2% fish gelatin, 150 mM NaCl, 0.1% Triton X-100, 0.1% sodium azide and 100 $\mu g\ ml^{-1}$ RNase A) for 1 h at 37 °C. Cells were incubated with the primary antibody (γH2AX) for 2 h, washed in PBS and incubated with secondary antibody for 1 h at room temperature. The secondary antibodies used were AlexaFluor 568 anti-IgG Mouse (Thermo Fisher Scientific, A-11004) or AlexaFluor 647 anti-IgG Mouse (Thermo Fisher Scientific, A-21235). The samples were finally fixed in 4% formaldehyde in PBS before fluorescence in situ hybridization (FISH) analysis.

Immunofluorescence

Cells were seeded onto glass coverslips 24 h before the experiment. Cells were fixed in 4% paraformadehyde in PBS for 10 min, washed in PBS and incubated in blocking solution of 5% BSA in PBS at room temperature. Cells were then incubated with primary antibodies for 2 h, washed in PBS and incubated with secondary antibodies for 1 h at room temperature. The secondary antibodies used were AlexaFluor 488 goat anti-rabbit IgG (H+L) (Thermo Fisher Scientific, A-11034) and AlexaFluor 568 goat anti-Mouse IgG (H+L) (Thermo Fisher Scientific, A-11004). The samples were finally washed in PBS and mounted in ProLong Diamond with DAPI (Invitrogen, P36971). For MitoTracker staining, MitoTracker Red CMXRos (Thermo Fisher Scientific, M7512) was added to the culture medium at a final concentration of 200 nM and incubated for 20 min under tissue culture conditions. After staining, cells were washed twice in prewarmed PBS, then fixed in paraformaldehyde 4% PBS for 10 min and washed in PBS. Imaging was performed using the Zeiss LSM 880 with the Airyscan microscope. ZEN (Zeiss) and ImageJ were used for image analysis.

Crystal violet assay for determining cell viability

Cells before crisis were seeded at low density (476 cells per cm^2) and kept at 37 °C for 10 days before fixing with 4% paraformaldehyde in PBS. Cells were then stained with 0.05% crystal violet in distilled water for 20 min.

fRIP

fRIP–seq was performed as previously described with minor modifications[61]. In brief, $IMR90^{E6E7}$ cells were cross-linked with 0.1% formaldehyde in PBS for 10 min at room temperature and unreacted formaldehyde was then neutralized with 125 mM glycine for 5 min at room temperature. Cells were washed twice with ice-cold PBS, collected by trypsinization and the cell pellets were resuspended in RIPA buffer (50 mM Tris pH 8, 150 mM KCl, 0.1% SDS, 1% Triton X-100, 5 mM EDTA, 0.5% sodium deoxycholate, 0.5 mM DTT (added fresh), protease inhibitor cocktail (Roche, 4693159001), 100 U ml^{-1} RNasin Ribonuclease Inhibitor (Promega, N251B)), and incubated for 20 min at 4 °C under slow rotation. The cell lysates were centrifuged at maximum speed at 4 °C for 10 min, and the supernatants were collected and diluted in equal volumes of freshly made

fRIP binding/wash buffer (150 mM KCl, 25 mM Tris pH 7.5, 5 mM EDTA, 0.5% NP-40, 0.5 mM DTT (added fresh), protease inhibitor cocktail (Roche, 4693159001), 100 U ml^{-1} RNasin Ribonuclease Inhibitor (Promega, N251B)). Diluted lysates were first precleared with Dynabeads Protein G (Invitrogen, 10004D) at a concentration of 25 µl of beads per 5 million cells for 30 min at 4 °C under slow rotation, and then incubated with 10 µg anti-Flag M2 antibodies (Sigma-Aldrich, F3165) previously coupled to 40 µl protein G Dynabeads (per each 5 million cells) for 2 h at 4 °C under slow rotation. The beads were washed twice with fRIP binding/wash buffer and protein–RNA cross-links were reversed by resuspending the beads in 56 µl of RNase-free water + 33 µl of 3× reverse-cross-linking buffer (3× PBS (without Ca^{2+} and Mg^{2+}), 6% *N*-lauroyl sarcosine, 30 mM EDTA, 15 mM DTT (added fresh), 10 µl of 20 mg ml^{-1} proteinase K (Millipore, 70663) and 1 µl of RNasin Ribonuclease Inhibitor. Protein degradation and reverse-cross-linking was performed for 1 h at 42 °C, followed by another hour at 55 °C. RNA was recovered by resuspending the beads and reaction buffer in 1 ml TRIzol Reagent (Invitrogen, 15596018) and purified using the Direct-zol RNA Microprep kit as recommended by manufacturer (Zymo Research, R2061).

Library generation and sequencing for fRIP–seq

RNA-seq libraries were prepared with immunoprecipitated RNA using the TruSeq stranded total RNA sample preparation kit according to the manufacturer's protocol (Illumina). RNA-seq libraries were multiplexed, normalized and pooled for sequencing. The libraries were sequenced on the MiniSeq system (Illumina) with paired-end 37 bp reads. Image analysis and base calling were performed using Illumina CASAVA v.1.8.2. on the MiniSeq system and sequenced reads were quality-tested using FASTQC. Read pairs were mapped individually to the most complete assembly available of human subtelomeres (http://www.wistar.org/lab/harold-c-riethman-phd/page/subtelomere-assemblies) using STAR v.2.5.3a allowing up to 101 mapping locations[62]. Primary alignments were assigned randomly, and each read is represented by one mapping location. Library sizes were normalized to 1×10^7 for comparison and the $\log_2$ ratio of enrichment of immunoprecipitates versus input was calculated at each base in the subtelomere using HOMER v.4.10 and the assembly available of human

subtelomeres[63]. Visualizations were performed using R v.3.6.1 (R Core Team). Graphical packages (Gviz v.1.28.3, rtracklayer v.1.44.2, gridExtra v.2.3, ggplot2 v.3.3.2) were used to visualize data.

RT–qPCR

Total RNA was isolated using TRIzol (Invitrogen, 15596018) and purified using the RNAeasy Mini Kit (Qiagen, 74106) according to the manufacturer's instructions. Genomic DNA was eliminated by double digestion with DNase I (RNase-free DNase Set, Qiagen, 79256). For RT–qPCR, 3.5 µg of RNA was reverse-transcribed either with random hexamers for measuring ISGs or with 2 µM of *telC*- and *GAPDH*-specific RT primers[64] using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, 18080-051). qPCR was performed on the CFX384 Touch Real-Time PCR Detection System (BioRad). Reactions were run in triplicates with Power SYBR Green Master Mix (Applied Biosystems, Thermo Fisher Scientific, 4367659) in a total volume of 10 µl with standard cycling conditions. Relative gene expression was normalized using *GAPDH* as a housekeeping gene and calculated using the Comparative CT Method ($\Delta\Delta C_T$ method). The primers are listed below.

RNA-dot blot

Total RNA (10 µg) was blotted onto a positively charged nitrocellulose membrane (GE, RPNBL/02/10). For fRIP, 200 ng of RNA from IP or input samples were used. For RNase controls, RNA was incubated with RNase A (Invitrogen, 12091-039) at 37 °C for 1 h. RNA was fixed by ultraviolet cross-linking (Stratalinker, 2400) and TERRA was detected by hybridizing overnight at 55 °C with a Church mix containing telomeric repeat probes generated by $[CCCTAA]^3$-primed Klenow labelling of an 800 bp $[TTAGGG]^n$ fragment in the presence of $[\alpha^{32}P]$dCTP. After hybridization, the membrane was washed twice in 2× SSC and 0.1% SDS for 10 min at room temperature and then once for 10 min at 50 °C. The radioactive signal was detected using the Typhoon FLA 9000 imager (GE Healthcare). After the signal acquisition, membranes were stripped and rehybridized at 50 °C overnight with ^{32}P-dCTP-labelled probes targeting *GAPDH* transcripts. Signals were measured using ImageJ.

TRF analysis

TRF analysis was performed as previously described[65]. In brief, genomic DNA was isolated by phenol–chloroform extraction and digested with AluI and MboI overnight. A total of 4 µg of digested gDNA was separated on 0.7% agarose gel at 40 V and transferred to a positively charged Nylon membrane (Amersham, RPN203B). After cross-linking the DNA and prehybridization (5× SSC. 0.1% *N*-lauroylsarcosine sodium salt solution, 0.04% SDS) for 2 h at 65 °C, the membrane was incubated with digoxigenin-labelled TelG probe diluted in hybridization buffer (1.3 nM final concentration) overnight at 65 °C. Digoxigenin-labelled TelG probe was generated as previously described[66]. Then, the membrane was washed three times with wash buffer 1 (2× SSC, 0.1% SDS), one time with wash buffer 2 (2× SSC) for 15 min each and blocked in freshly prepared blocking solution (100 mM maleic acid, 150 mM NaCl, pH 7.5, 1% (w/v) blocking reagent (Roche, 11096176001)) for 30 min. Next, the membrane was incubated for 30 min in anti-digoxigenin-AP antibodies (Roche, 11093274910) diluted in blocking solution, washed twice in wash buffer 3 (100 mM maleic acid, 150 mM NaCl, pH 7.5, 0.3% (v/v) Tween-20) for 15 min each and equilibrated in AP buffer (100 mM Tris, 100 mM NaCl, pH 9.5) for 2 min. Digoxigenin-labelled telomeric DNA was detected using CDP-star ready to use (Roche, 12041677001) solution.

Genome-wide CRISPR screen

The human Brunello CRISPR knockout pooled library was obtained from D. Root and J. Doench through Addgene (73179-LV)[67]. This ready-to-use lentiviral library has 77,441 gRNAs, targeting 19,114 protein-coding genes, with approximately 4 sgRNAs per gene. For adequate representation of each sgRNA, a total of 100 million pre-crisis (PD100) IMR90^{E6E7} fibroblasts were transduced with the lentiviral library at a multiplicity of infection of 0.5. Transductions were performed in six-well plates (3 million cells per well) in medium containing 4 µg ml^{-1} polybrene while centrifuging at 1,000 rcf for 1 h at 33 °C (spinfection). The next day, cells were transferred to Cell Factory System (Thermo Fisher Scientific, 140360), and puromycin-containing medium (1 µg ml^{-1}) was added for 7 days to eliminate uninfected cells and achieve genome-edited cell pools. After selection, cells were

pooled together and divided into two technical replicates of 30 million cells each corresponding to the library baseline control at day 0. Another 30 million cells were replated into one Cell Factory System and positive selection was performed by growing cells for additional 15 days, at which point crisis-associated cell death was frequent. Two technical replicates of 30 million cells each were prepared by collecting cells at day 15, and their genomic DNA was extracted using a modified version of QIAGEN's DNeasy Blood and Tissue Kit provided by the FLI-Seq Library Prep for CRISPR kit (Eclipse Bioinnovations). Genomic DNA was also extracted from the baseline count control sample at day 0. DNA fragments containing the sgRNA sequences were first captured from sheared gDNA, amplified by PCR using the FLI-Seq Library Prep for CRISPR kit (Eclipse Bioinnovations) and processed for next-generation sequencing. CRISPR libraries were multiplexed, normalized and pooled for sequencing. To compensate for low base diversity in CRISPR libraries, high-diversity libraries or PhiX Control v3 Library were spiked in for sequencing and the libraries were sequenced on the HiSeq 2500 system (Illumina) as single reads. Image analysis and base calling were performed using Illumina CASAVA v.1.8.2 on the HiSeq 2500 system and sequenced reads were quality-tested using FASTQC. Fold changes of sgRNA read counts before (day 0) and after (day 15) enrichment were calculated with the PinAPL-Py software and hits were defined as genes with a read count ratio in the survival pool to the baseline control $\log_2$[fold change] ≥ 2 in both replicates. GO enrichment analysis was performed using the WEB-based Gene Set Analysis Toolkit (WebGestalt). The top 20 GO terms with an FDR value of <0.05 were considered to be statistically significant and visualized using the ClueGO and ggplot2 R packages.

Whole-transcriptome analysis

Total RNA was isolated using TRIzol (Invitrogen, 15596018) and purified using the RNAeasy Mini Kit (Qiagen, 74106) according to the manufacturer's instructions. Genomic DNA was eliminated by double digestion with DNase I (RNase-free DNase Set, Qiagen, 79256). The quality of the isolated total RNA was assessed using the Agilent TapeStation 4200 and RNA-seq libraries were prepared with 500 ng total RNA using the TruSeq stranded mRNA sample preparation kit according to the

manufacturer's protocol (Illumina). RNA-seq libraries were multiplexed, normalized and pooled for sequencing. The libraries were sequenced on the HiSeq 4000 system (Illumina) as 50 bp single reads or the NextSeq 500 system (Illumina) as 75 bp single reads. Image analysis and base calling were performed using Illumina CASAVA v.1.8.2 on the HiSeq 4000 system and sequenced reads were quality-tested using FASTQC. Sequenced reads were quality-tested using FASTQC (v.0.11.8)[68] and mapped to the hg19 human genome using the STAR aligner (v.2.5.3a)[62] with the default parameters. Raw or transcripts per kilobase million (TPM) gene expression was quantified across all the exons of RefSeq genes with analyzeRepeats.pl in HOMER (v.4.11.1)[63], which used the top-expressed isoform as proxy for gene expression. Differential gene expression was performed on the raw gene counts with the R package, DESeq2 (v.1.24.0)[69], using replicates to compute within-group dispersion. Differentially expressed genes were defined as having a FDR < 0.05 and a $\log_2$[fold change] > 0.585 (~1.5 fold) when comparing two experimental conditions. GO enrichment analysis was performed using the WEB-based Gene Set Analysis Toolkit" (WebGestalt) and the R package clusterProfiler. The top 20 GO terms with an FDR value of <0.05 were considered to be statistically significant and visualized using the ClueGO and ggplot2 R packages. Heat maps and Venn diagrams were generated using the R packages ComplexHeatmap and VennDiagram, respectively.

ELISA assay

One million cells were seeded in a 10 cm plate for 24 h. Conditioned cell culture medium was collected and centrifuged at 14,000 rpm for 10 min at 4 °C to remove cellular debris. Three technical replicates of 50 µl each were collected for the assay. For preparation of standards, GlutaMax-DMEM (Gibco, 10569-010) supplemented with 0.1 mM non-essential amino acids (Corning, 25-025-Cl) and 15% fetal bovine serum (Thermo Fisher Scientific, SH3007103) was used. IFNβ secretion in cell culture supernatant was quantified using the IFNβ enzyme-linked immunosorbent assay (ELISA) kit (R&D Systems Human IFNβ Quantikine ELISA Kit, DIFNB0) according to the manufacturer's instruction. Supernatants from the treated cells were collected and incubated in IFNβ ELISA kit for 2 h at room temperature and washed three times with wash buffer. The optical density at

450 nm for each sample was measured with a microplate reader and data obtained were plotted against a four-parameter logistic standard curve to determine the concentration of IFNβ.

Chemical reagents

The reagents were as follows: recombinant human IFN-beta protein: R&D Systems, 8499-IF-010/CF; poly(dA:dT): InvivoGen, tlrl-patn; poly(I:C): InvivoGen, tlrl-pic; 2′,3′-dideoxycytidine: Sigma-Aldrich, D5782; deoxyribonucleic acid sodium salt from herring testes (HT-DNA): Sigma-Aldrich, D6898; MitoTracker: Invitrogen, M22426; IncuCyte Nuclight Rapid red dye: Sartorius, 4717; IncuCyte Cytotox red dye: Sartorius, 4632; IncuCyte Cytotox green dye: Sartorius, 4633; doxycycline: Sigma-Aldrich, D9891; staurosporine: Cell Signaling Technology, 9953; recombinant human TNF-alpha protein: R&D Systems, 210-TA; BV6: Selleckchem, S7597; z-VAD-FMK: Cell Signaling Technology, 60332; LPS: Cell Signaling Technology, 14011; nigericin: Cayman Chemical, 11437; Hoechst 33342: Thermo Fisher Scientific, 62249.

Statistical analysis

Statistical analysis was performed using Prism 9. Comparisons between two groups were performed using unpaired two-tailed Student's *t*-tests. Multiple comparisons were performed using one-way ANOVA followed by Tukey's or Dunnett's multiple-comparisons test. For all representative findings, two or three independent experiments were performed, and similar results were obtained. Significance in all figures is denoted as follows: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Exact *P* values are as follows.

Figure 1c: 36 h after incubation with Cytotox. Vector versus WT (<0.0001); vector versus mRHIM2 (<0.0001).

Figure 1d: IFNB1: vector versus WT (<0.0001); vector versus mZα2 (0.9775); vector versus mRHIM1 (0.9554); vector versus mRHIM2 (<0.0001). CCL5: vector versus WT (<0.0001); vector versus mZα2 (0.101); vector versus mRHIM1 (0.9799); vector versus mRHIM2 (<0.0001).

CXCL10: vector versus WT (<0.0001); vector versus mZα2 (0.8118); vector versus mRHIM1 (0.9958); vector versus mRHIM2 (<0.0001). IFI44: vector versus WT (<0.0001); vector versus mZα2 (0.9426); vector versus mRHIM1 (0.9997); vector versus mRHIM2 (<0.0001). ISG15-vector versus WT (<0.0001); vector versus mZα2 (0.1726); vector versus mRHIM1 (0.8483); vector versus mRHIM2 (<0.0001).

Figure 2b: IFNB1: shScramble versus sh*TRF2* (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 1 (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 2 (<0.0001). CCL5: shScramble versus sh*TRF2* (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 1 (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 2 (<0.0001). CXCL10: shScramble versus sh*TRF2* (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 1 (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 2 (<0.0001). IFI44: shScramble versus sh*TRF2* (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 1 (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 2 (<0.0001). ISG15-shScramble versus sh*TRF2* (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 1 (<0.0001); sh*TRF2*-siCtrl versus sh*TRF2*-si*ZBP1* 2 (<0.0001).

Figure 3a: 9p, growing versus crisis (<0.0001). 13q, growing versus crisis (<0.0001). 15q, growing versus crisis (<0.0001). Xp, growing versus crisis (<0.0001). Xq, growing versus crisis (<0.0001). Figure 3b: 9p, shScramble versus sh*TRF2* (<0.0001). 13q, shScramble versus sh*TRF2* (<0.0001). 15q, shScramble versus sh*TRF2* (<0.0001). Xp, shScramble versus sh*TRF2* (<0.0001). Xq, shScramble versus sh*TRF2* (0.0015).

Figure 3f: 28 h after incubation with Cytotox. ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001).

Figure 4a: PD45 versus PD79 (<0.0001); PD45 versus PD99 (<0.0001); PD45 versus PD103 (<0.0001); PD45 versus PD106 (<0.0001); PD35 versus PD50 (0.0043); PD35 versus PD62 (<0.0001); PD35 versus PD71 (<0.0001); PD35 versus PD86 (<0.0001).

Figure 4c: CCL5: PD40 vector versus PD40 WT (>0.9999). PD40 vector versus PD40 mZα (>0.9999); PD40 vector versus PD40 mRHIM1 (>0.9999); PD40 vector versus PD40 mRHIM2 (>0.9999); PD95 vector versus PD95 WT (<0.0001); PD95 vector versus PD95 mZα (0.8488); PD95

vector versus PD95 mRHIM1 (0.9998); PD95 vector versus PD95 mRHIM2 (<0.0001). CXCL10: PD40 vector versus PD40 WT (>0.9999); PD40 vector versus PD40 mZα (>0.9999); PD40 vector versus PD40 mRHIM1 (>0.9999); PD40 vector versus PD40 mRHIM2 (>0.9999); PD95 vector versus PD95 WT (0.0005); PD95 vector versus PD95 mZα (0.9766); PD95 vector versus PD95 mRHIM1 (>0.9999); PD95 vector versus PD95 mRHIM2 (<0.0001). IFI44: PD40 vector versus PD40 WT (>0.9999); PD40 vector versus PD40 mZα (>0.9999); PD40 vector versus PD40 mRHIM1 (0.9965); PD40 vector versus PD40 mRHIM2 (>0.9999); PD95 vector versus PD95 WT (<0.0001); PD95 vector versus PD95 mZα (>0.9566); PD95 vector versus PD95 mRHIM1 (0.0338); PD95 vector versus PD95 mRHIM2 (<0.0001).

Figure 4e: CCL5: Flag–ZBP1(S)–FIS1 versus Flag–ZBP1(S) (<0.0001); vector versus Flag–ZBP1(S)–FIS1 (<0.0001). CXCL10: Flag–ZBP1(S)–FIS1 versus Flag-ZBP1(S) (<0.0001); vector versus Flag–ZBP1(S)–FIS1 (<0.0001). IFI44: Flag–ZBP1(S)–FIS1 versus Flag–ZBP1(S) (<0.0001); vector versus Flag–ZBP1(S)–FIS1 (<0.0001).

Figure 4g: CCL5: vector siCtrl versus vector si*MAVS* (0.9002); Flag–ZBP1(S) siCtrl versus Flag–ZBP1(S) si*MAVS* (0.3629); Flag–ZBP1(S)–FIS1 siCtrl versus Flag–ZBP1(S)–FIS1 si*MAVS* (<0.0001). CXCL10: vector siCtrl versus vector si*MAVS* (0.9537). Flag–ZBP1(S) siCtrl versus Flag–ZBP1(S) si*MAVS* (0.0001). Flag–ZBP1(S)–FIS1 siCtrl versus Flag–ZBP1(S)–FIS1 si*MAVS* (<0.0001). IFI44: vector siCtrl versus vector si*MAVS* (>0.9999). Flag–ZBP1(S) siCtrl versus Flag–ZBP1(S) si*MAVS* (0.9966). Flag–ZBP1(S)–FIS1 siCtrl versus Flag–ZBP1(S)–FIS1 si*MAVS* (<0.0001).

Extended Data Fig. 3d: IMR90^{E6E7}, 24 h after incubation with Cytotox. sgGFP versus sgZBP1 1 (<0.0001). sgGFP versus sgZBP1 2 (<0.0001); sgGFP versus sgZBP1 3 (<0.0001); sgLUC versus sgZBP1 1 (<0.0001); sgLUC versus sgZBP1 2 (<0.0001); sgLUC versus sgZBP1 3 (<0.0001). WI38^{SV40LT}, 24 h after incubation with Cytotox. sgGFP versus sgZBP1 1 (<0.0001); sgGFP versus sgZBP1 2 (<0.0001); sgGFP versus sgZBP1 3 (<0.0001); sgLUC versus sgZBP1 1 (<0.0001); sgLUC versus sgZBP1 2 (<0.0001); sgLUC versus sgZBP1 3 (<0.0001).

Extended Data Fig. 3e: IMR90^{E6E7}: sgLUC versus sgGFP (0.0668); sgLUC versus sgZBP1 1 (<0.0001); sgLUC versus sgZBP1 2 (<0.0001); sgLUC versus sgZBP1 3 (<0.0001). WI38^{SV40LT}: sgLUC versus sgGFP (0.1293); sgLUC versus sgZBP1 1 (<0.0001); sgLUC versus sgZBP1 2 (<0.0001); sgLUC versus sgZBP1 3 (<0.0001).

Extended Data Fig. 3f: IMR90^{E6E7}: sgLUC versus sgGFP (0.9995); sgLUC versus sgZBP1 1 (0.0002); sgLUC versus sgZBP1 2 (<0.0001); sgLUC versus sgZBP1 3 (0.0004). WI38^{SV40LT}: sgLUC versus sgGFP (0.7845); sgLUC versus sgZBP1 1 (0.0018); sgLUC versus sgZBP1 2 (<0.0001). sgLUC versus sgZBP1 3 (0.0011).

Extended Data Fig. 4b: IMR90^{E6E7}: growing versus crisis (<0.0001). WI38^{SV40LT}: growing versus crisis (<0.0001).

Extended Data Fig. 4e: experiment 1: siCtrl versus si*IFNAR1* (1.39178×10^{-5}); siCtrl versus si*IFNAR2* (8.05855×10^{-5}); siCtrl versus si*CGAS* (3.65474×10^{-5}); siCtrl versus si*AIM2* (0.57384863); siCtrl versus si*TLR3* (0.028940902); siCtrl versus si*TLR7* (0.130633803); siCtrl versus si*TLR9* (0.481604543); siCtrl versus si*RIG-I* (0.067963804); siCtrl versus si*MDA5* (0.078645098); siCtrl versus RIG-I + si*MDA5* (0.075426998); siCtrl versus si*STING* (0.000445511); siCtrl versus si*PYCARD* (0.272857261); siCtrl versus si*MyD88* (0.16111072); siCtrl versus si*TRIF* (0.138938302); siCtrl versus si*MAVS* (3.61752×10^{-5}). Experiment 2: siCtrl versus si*IFNAR1* (1.48978×10^{-6}); siCtrl versus si*IFNAR2* (1.4765×10^{-5}); siCtrl versus si*CGAS* (1.40243×10^{-6}); siCtrl versus si*AIM2* (0.081532291); siCtrl versus si*TLR3* (0.244271443); siCtrl versus si*TLR7* (0.413559022); siCtrl versus si*TLR9* (0.1970414); siCtrl versus si*RIG-I* (0.015458711); siCtrl versus si*MDA5* (0.027374775); siCtrl versus RIG-I + si*MDA5* (0.857465154); siCtrl versus si*STING* (9.9925×10^{-5}); siCtrl versus si*PYCARD* (0.28638452); siCtrl versus si*MyD88* (0.73179998); siCtrl versus si*TRIF* (0.0188581); siCtrl versus si*MAVS* (7.6601×10^{-6}). Experiment 3: siCtrl versus si*IFNAR1* (1.23834×10^{-5}); siCtrl versus si*IFNAR2* (6.7646×10^{-5}); siCtrl versus si*CGAS* (1.04549×10^{-6}); siCtrl versus si*AIM2* (0.05401101); siCtrl versus si*TLR3* (0.16170163); siCtrl versus si*TLR7* (0.01702073); siCtrl versus si*TLR9* (0.0189054); siCtrl versus si*RIG-I* (0.00405314); siCtrl versus si*MDA5* (0.373743841); siCtrl versus RIG-I + si*MDA5*

(0.024471461); siCtrl versus si*STING* (0.00368923); siCtrl versus si*PYCARD* (0.08524613); siCtrl versus si*MyD88* (0.14064815); siCtrl versus si*TRIF* (0.12424426); siCtrl versus si*MAVS* (0.008386476).

Extended Data Fig. 5d: IFNB1: vector versus WT (<0.0001); vector versus ΔZα2 (0.2671); vector versus ΔZα2ΔRHIM1ΔRHIM2 (0.1795); vector versus ΔC (<0.0001); vector versus ΔRHIM1 (0.6671); vector versus ΔRHIM2 (<0.0001); vector versus ΔRHIM1ΔRHIM2 (0.0151). CCL5: vector versus WT (<0.0001); vector versus ΔZα2 (0.0367); vector versus ΔZα2ΔRHIM1ΔRHIM2 (0.978); vector versus ΔC (<0.0001); vector versus ΔRHIM1 (0.0675); vector versus ΔRHIM2 (<0.0001); vector versus ΔRHIM1ΔRHIM2 (0.0816). CXCL10: vector versus WT (<0.0001); vector versus ΔZα2 (0.1945); vector versus ΔZα2ΔRHIM1ΔRHIM2 (0.9998); vector versus ΔC (<0.0001); vector versus ΔRHIM1 (0.1365); vector versus ΔRHIM2 (<0.0001); vector versus ΔRHIM1ΔRHIM2 (0.1988). IFI44: vector versus WT (<0.0001); vector versus ΔZα2 (0.0294); vector versus ΔZα2ΔRHIM1ΔRHIM2 (0.9813); vector versus ΔC (<0.0001); vector versus ΔRHIM1 (0.0287); vector versus ΔRHIM2 (<0.0001); vector versus ΔRHIM1ΔRHIM2 (0.2269). ISG15: vector versus WT (<0.0001); vector versus ΔZα2 (0.0003); vector versus ΔZα2ΔRHIM1ΔRHIM2 (0.0242); vector versus ΔC (<0.0001); vector versus ΔRHIM1 (0.0005); vector versus ΔRHIM2 (<0.0001); vector versus ΔRHIM1ΔRHIM2 (0.0001).

Extended Data Fig. 6e: vector versus ZBP1(S)–Flag (0.0082); siCtrl versus si*ZBP1* (<0.0001); siCtrl versus si*RIPK3* (0.9942); siCtrl versus siMLK (>0.9999); siCtrl versus siRIPK1 (>0.9999); siCtrl versus siFADD (0.9425); siCtrl versus siCASP8 (>0.9999); siCtrl versus siNLRP3 (>0.9999); siCtrl versus siPYCARD (>0.9999); siCtrl versus siCASP1 (0.9790); siCtrl versus siIFNAR1 (<0.0001); siCtrl versus siIFNAR2 (<0.0001); siCtrl versus siSTAT1 (<0.0001); siCtrl versus siSTAT2 (<0.0001); siCtrl versus siIRF9 (<0.0001); siCtrl versus siATG5 (<0.0001); siCtrl versus siATG7 (<0.0001); siCtrl versus siATG12 (<0.0001).

Extended Data Fig. 6g: mock versus shScramble day 3 (0.9934); mock versus shScramble day 6 (0.9996); mock versus shScramble day 9 (0.9994); mock versus shScramble day 12 (0.9996); mock versus sh*TRF2* day 3 (<0.0001); mock versus sh*TRF2* day 6 (<0.0001); mock versus sh*TRF2* day 9 (<0.0001); mock versus sh*TRF2* day 12 (<0.0001).

Extended Data Figure 6h: mock versus shScramble day 3 (>0.9999); mock versus shScramble day 6 (>0.9999); mock versus shScramble day 9 (>0.9999); mock versus shScramble day 12 (>0.9999); mock versus sh*TRF2* day 3 (<0.0001); mock versus sh*TRF2* day 6 (<0.0001); mock versus sh*TRF2* day 9 (<0.0001); mock versus sh*TRF2* day 12 (<0.0001).

Extended Data Figure 6j: siCtrl versus si*ZBP1* 1 (0.9038); siCtrl versus si*ZBP1* 2 (0.2923).

Extended Data Fig. 7c: IMR90$^{\mathrm{E6E7}}$: PD42 versus PD45 (>0.9999); PD42 versus PD46 (0.9999); PD42 versus PD51 (0.9999); PD42 versus PD55 (0.9997); PD42 versus PD58 (>0.9999); PD42 versus PD62 (0.9998); PD42 versus PD69 (0.9996); PD42 versus PD73 (>0.9999); PD42 versus PD85 (0.9996); PD42 versus PD87 (0.9998); PD42 versus PD90 (0.9998); PD42 versus PD99 (0.9864); PD42 versus PD100 (0.0891); PD42 versus PD101 (0.0003); PD42 versus PD103 (<0.0001); PD42 versus PD105 (<0.0001); PD42 versus PD106 (<0.0001). IMR90$^{\mathrm{E6E7}}$, hTERT: PD61 versus PD74 (0.8616); PD61 versus PD89 (0.9996); PD61 versus PD95 (0.9995); PD61 versus PD105 (>0.9999); PD61 versus PD108 (0.9882). WI38$^{\mathrm{SV40LT}}$: PD35 versus PD38 (>0.9999); PD35 versus PD41 (>0.9999); PD35 versus PD44 (0.9997); PD35 versus PD47 (>0.9999); PD35 versus PD50 (0.9999); PD35 versus PD53 (>0.9999); PD35 versus PD58 (0.9998); PD35 versus PD62 (0.9998); PD35 versus PD67 (0.9998); PD35 versus PD69 (0.9994); PD35 versus PD71 (0.9997); PD35 versus PD79 (0.9994). PD35 versus PD82 (0.9990); PD42 versus PD86 (<0.0001); PD42 versus PD87 (<0.0001); PD42 versus PD89 (<0.0001); PD42 versus PD90 (<0.0001). WI38$^{\mathrm{SV40LT}}$, hTERT: PD50 versus PD63 (0.9576); PD50 versus PD74 (0.9576); PD50 versus PD77 (0.9796); PD50 versus PD83 (>0.9999); PD50 versus PD92 (0.9576).

Extended Data Fig. 8a: growing versus crisis (<0.0001).

Extended Data Fig. 8b: shScramble versus sh*TRF2* (<0.0001).

Extended Data Fig. 9c: 9p: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (<0.0001). 13q: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN)

(0.0002); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (0.0005). 15q: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (<0.0001). Xp: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (<0.0001). Xq: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (0.0023); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (0.0009).

Extended Data Figure 10b: CCL5: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (0.7919). CXCL10–ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (0.9996). IFI44: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(ΔZα2)–Flag TRF1(ΔN) versus ZBP1(S)(ΔZα2)–Flag VP16–TRF1(ΔN) (>0.9999).

Extended Data Figure 10d: CCL5: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001); ZBP1(S)(mZα2)–Flag TRF1(ΔN) versus ZBP1(S)(mZα2)–Flag VP16–TRF1(ΔN) (0.8595). CXCL10: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001). ZBP1(S)mZα2–Flag TRF1(ΔN) versus ZBP1(S)(mZα2)–Flag VP16–TRF1(ΔN) (>0.9999). IFI44: ZBP1(S)(WT)–Flag TRF1(ΔN) versus ZBP1(S)(WT)–Flag VP16–TRF1(ΔN) (<0.0001). ZBP1(S)(mZα2)–Flag TRF1(ΔN) versus ZBP1(S)(mZα2)–Flag VP16–TRF1(ΔN) (0.9998).

Extended Data Figure 11c: IMR90^{E6E7}: TOM20–ZBP1, growing versus crisis (<0.0001); MitoTracker–ZBP1, growing versus crisis (<0.0001); TFAM–ZBP1, growing versus crisis (<0.0001). WI38^{SV40LT}: TOM20–ZBP1, growing versus crisis (<0.0001); MitoTracker–ZBP1, growing versus crisis (<0.0001); TFAM–ZBP1, growing versus crisis (<0.0001).

Extended Data Figure 11d: dsDNA intensity: growing versus crisis (<0.0001); dsRNA intensity: growing versus crisis (<0.0001); number of ZBP1 filaments: growing versus crisis (0.0712).

Extended Data Figure 11e: dsRNA intensity: siCtrl versus si*SUV3* (<0.0001); siCtrl versus si*PNPase* (<0.0001); number of ZBP1 filaments: siCtrl versus si*SUV3* (0.6179). siCtrl versus si*PNPase* (0.3240).

Extended Data Figure 12c: CCL5: ZBP1(S)(WT)–Flag TRF1(ΔN): siCtrl versus si*MAVS* (0.6383); siCtrl versus si*MDA5* (0.9499); siCtrl versus si*RIG-I* (>0.9999); siCtrl versus si*MDA5* + *RIG-I* (0.9959). ZBP1(S)(WT)–Flag VP16–TRF1(ΔN): siCtrl versus si*MAVS* (<0.0001), siCtrl versus si*MDA5* (0.2942); siCtrl versus si*RIG-I* (0.9933); siCtrl versus si*MDA5* + *RIG-I* (0.8552). CXCL10: ZBP1(S)(WT)–Flag TRF1(ΔN): siCtrl versus si*MAVS* (0.9997); siCtrl versus si*MDA5* (0.9980); siCtrl versus si*RIG-I* (0.9973); siCtrl versus si*MDA5* + *RIG-I* (>0.9999). ZBP1(S)(WT)–Flag VP16–TRF1(ΔN): siCtrl versus si*MAVS* (<0.0001); siCtrl versus si*MDA5* (0.0645); siCtrl versus si*RIG-I* (0.8284); siCtrl versus si*MDA5* + *RIG-I* (0.1366). IFI44: ZBP1(S)(WT)–Flag TRF1(ΔN): siCtrl versus si*MAVS* (0.9999); siCtrl versus si*MDA5* (>0.9999); siCtrl versus si*RIG-I* (>0.9999); siCtrl versus si*MDA5* + *RIG-I* (>0.9999). ZBP1(S)(WT)–Flag VP16–TRF1(ΔN): siCtrl versus si*MAVS* (<0.0001); siCtrl versus si*MDA5* (<0.0001); siCtrl versus si*RIG-I* (<0.0001); siCtrl versus si*MDA5* + *RIG-I* (<0.0001).

Extended Data Figure 13d: IMR90$^{\mathrm{E6E7}}$: vector siCtrl versus si*MAVS* (>0.9999); vector siCtrl versus si*IFNAR2* (>0.9999); Flag–ZBP1(S)–FIS1 siCtrl versus si*MAVS* (<0.0001); Flag–ZBP1(S)–FIS1 siCtrl versus si*IFNAR2* (<0.0001). HMECs: vector siCtrl versus si*MAVS* (>0.9999); vector siCtrl versus si*IFNAR2* (>0.9999); Flag–ZBP1(S)–FIS1 siCtrl versus si*MAVS* (<0.0001); Flag–ZBP1(S)–FIS1 siCtrl versus si*IFNAR2* (<0.0001).

shRNAs

shRNAs were as follows: Scramble, CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG, pLKO.1, Addgene, 1864; *TRF2*, ACAGAAGCAGTGGTCGAATC, pLKO.1, Open Biosystems, TRCN0000018358; sh*ZBP1* 1,

CCAAGTCCTCTACCGAATGAA, pLKO.1; sh*ZBP1* 2, GCACAATCCAATCAACATGAT, pLKO.1, Sigma-Aldrich, TRCN0000123050.

siRNAs

siRNAs were as follows: non-targeting pool (Dharmacon, D-001810-10-20): UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUACAUGUUUUCCUA); non-targeting (Dharmacon, D-001810-01-05): UGGUUUACAUGUCGACUAA; *ZBP1* 1 (Dharmacon, J-014650-07-0005): GGACACGGGAACAUCAUUA; *ZBP1* 2 (Dharmacon, J-014650-08-0005): CAAAAGAUGUGAACCGAGA; *ZBP1* 3 (Dharmacon, J-014650-06-0005): GGAUUUCCAUUGCAAACUC; *ZBP1* 4 (Dharmacon, J-014650-05-0005): CAAAGUCAGCCUCAAUUAU; *MB21D1* (Dharmacon, L-015607-02-0005): GAAGAAACAUGGCGGCUAU, AGGAAGCAACUACGACUAA, AGAACUAGAGUCACCCUAA, CCAAGAAGGCCUGCGCAUU; *TMEM173* (Dharmacon, L-024333-02-0005): UCAUAAACUUUGGAUGCUA, CGAACUCUCUCAAUGGUAU, AGCUGGGACUGCUGUUAAA, GCAGAUGACAGCAGCUUCU; *MAVS* 1 (Dharmacon, J-024237-07-0005): GCAAUGUGGAUGUUGUAGA; *MAVS* 2 (Dharmacon, J-024237-05-0005): AAGUAUAUCUGCCGCAAUU; *MAVS* 3 (Dharmacon, J-024237-06-0005): CAUCCAAAGUGCCUACUAG; *MAVS* 4 (Dharmacon, J-024237-08-0005): CAUCCAAAUUGCCCAUCAA; *PYCARD* (Dharmacon, L-004378-00-0005): GGAAGGUCCUGACGGAUGA, UCACAAACGUUGAGUGGCU, GGCCUGCACUUUAUAGACC, CCACCAACCCAAGCAAGAU; *MYD88* (Dharmacon, L-004769-00-0005): CGACUGAAGUUGUGUGUGU, GCUAGUGAGCUCAUCGAAA, GCAUAUGCCUGAGCGUUUC, GCACCUGUGUCUGGUCUAU; *TLR7* (Dharmacon, L-004714-00-0005): CAACAACCGGCUUGAUUUA, GGAAAUUGCCCUCGUUGUU, GAAUCUAUCACAAGCAUUU, GGAAUUACUCAUAUGCUAA; *TLR3* (Dharmacon, L-007745-00-0005): GAACUAAAGAUCAUCGAUU, CAGCAUCUGUCUUUAAUAA, AGACCAAUCUCUCAAAUUU, UCACGCAAUUGGAAGAUUA; *TLR9* (Dharmacon, L-004066-00-0005): CAGACUGGGUGUACAACGA, GCAAUGCACUGGGCCAUAU, CGGCAACUGUUAUUACAAG, ACAAUAAGCUGGACCUCUA; *TLR8* (Dharmacon, L-4715-00-0005): CAACGGAAAUCCCGGUAUA, CAGAAUAGCAGGCGUAACA, GUGCAGCAAUCGUCGACUA,

CUUCCAAACUUAUCGACUA; *DDX58* (Dharmacon, L-012511-00-0005): GCACAGAAGUGUAUAUUGG, CCACAACACUAGUAAACAA, CGGAUUAGCGACAAAUUUA, UCGAUGAGAUUGAGCAAGA; *IFIH1* (Dharmacon, L-013041-00-0005): GAAUAACCCAUCACUAAUA, GCACGAGGAAUAAUCUUUA, UGACACAAUUCGAAUGAUA, CAAUGAGGCCCUACAAAUU; *TICAM1* (Dharmacon, L-012833-00-0005): GGAGCCACAUGUCAUUUGG, CCAUAGACCACUCAGCUUU, GGACGAACACUCCCAGAUC, CCACUGGCCUCCCUGAUAC; *IFNAR1* (Dharmacon, L-020209-00-0005): GCGAAAGUCUUCUUGAGAU, UGAAACCACUGACUGUAUA, GAAAAUUGGUGUCUAUAGU, GAAGAUAAGGCAAUAGUGA; *IFNAR2* (Dharmacon, L-015411-00-0005): CAGAGGGAAUUTUUAAGAA, GAGUAAACCAGAAGAUUUG, CACCAGAGUUUGAGAUUGU, UCACCUAUAUCAUUGACAA; *AIM2* (Dharmacon, L-011951-00-0005): GCACAGUGGUUUCUUAGAG, UCAGACGAGUUUAAUAUUG, GAAAGUUGAUAAGCAAUAC, GUUCAUAGCACCAUAAAGG; *SUPV3L1* (Dharmacon, L-017841-01-0005): UGGCUAAGCUACCGAUUUA, GUAAGGAUGAUCUACGUAA, CGGUGCAGCUCAUGCGGAU, GGAAAGACUUAUCACGCAA; *PNPT1* (Dharmacon, L-020112-00-0005): GACAGAAGUAGUAUUGUAA, ACAGAAAGAUUAUUGGCUA, GAAUGUAAGUUGUGAGGUA, AAUCAGAGAUACUGGUGUA; *RIPK3* (Dharmacon, L-003534-00-0005): CCACAGGGUUGGUAUAAUC, AACCAGCACUCUCGUAAUG, GCUACGAUGUGGCGGUCAA, GACCGCUCGUUAACAUAUA; *ATG7* (Dharmacon, L-020112-00-0005): CCAACACACUCGAGUCUUU, GAUCUAAAUCUCAAACUGA, GCCCACAGAUGGAGUAGCA, GCCAGAGGAUUCAACAUGA; TERRA 1 (Dharmacon, CTM-536949): AGGGUUAGGGUUAGGGUUAUU; TERRA 2 (Dharmacon, CTM-536950): GGGUUAGGGUUAGGGUUAGUU; *MLKL* (Dharmacon, L-005326-00-0005): GAGCAACGCAUGCCUGUUU, CAAACUUCCUGGUAACUCA, GAAGGAGCUCUCGCUGUUA, GGAUUUGCAUUGAUGAAAC; *RIPK1* (Dharmacon, L-004445-00-0005): CCACUAGUCUGACGGAUAA, UGAAUGACGUCAACGCAAA, GCACAAAUACGAACUUCAA, GAUGAAAUCCAGUGACUUC); *FADD* (Dharmacon, L-003800-00-

0005): CAUUUAACGUCAUAUGUGA, GGAGAAGGCUGGCUCGUCA, UGACAGAGCGUGUGCGGGA, GCAUCUACCUCCGAAGCGU); *CASP8* (Dharmacon, L-003466-00-0005): GGACAAAGUUUACCAAAUG, GCCCAAACUUCACAGCAUU, GAUAAUCAACGACUAUGAA, GUCAUGCUCUAUCAGAUUU); *NLRP3* (Dharmacon, L-017367-00-0005): GGAUCAAACUACUCUGUGA, UGCAAGAUCUCUCAGCAAA, GAAGUGGGGUUCAGAUAAU, GCAAGACCAAGACGUGUGA); *CASP1* (Dharmacon, L-004401-00-0005): GGAAGACUCAUUGAAACAUA, GAUGGUAGAGCGCAGAUGC, CCGCAAGGUUCGAUUUUCA, GAGUGACUUUGACAAGAUG); *STAT1* (Dharmacon, L-003543-00-0005): GCACGAUGGGCUCAGCUUU, CUACGAACAUGACCCUAUC, GAACCUGACUUCCAUGCGG, AGAAAGAGCUUGACAGUAA); *STAT2* (Dharmacon, L-012064-00-0005): GGACUGAGUUGCCUGGUUA, GGACUGAGGAUCCAUUAUU, GAGCCCUCCCUGGCAAGUUA, GAUUUGCCCUGUGAUCUGA); *IRF9* (Dharmacon, L-020858-00-0005): GCAGAGACUUGGUCAGGUA, CCACCGAAGUUCCAGGUAA, GCGUGGAGCUCUUCAGAAC, GAAAGUACCAUCAAAGCGA); *ATG5* (Dharmacon, L-004374-00-0005): GGCAUUAUCCAAUUGGUUU, GCAGAACCAUACUAUUUGC, UGACAGAUUUGACCAGUUU, ACAAAGAUGUGCUUCGAGA); *ATG12* (Dharmacon, L-010212-00-0005): GAACACCAAGUUUCACUGU, GCAGUAGAGCGAACACGAA, GGGAAGGACUUACGGAUGU, GGGAUGAACCACAAAGAAA).

Primers for RT–qPCR

The primers were as follows: 9p-F, GAGATTCTCCCAAGGCAAGG; 9p-R, ACATGAGGAATGTGGGTGTTAT; 13q-F, CTGCCTGCCTTTGGGATAA; 13q-R, AAACCGTTCTAACTGGTCTCTG; 15q-2-F, CAGCGAGATTCTCCCAAGCTAAG; 15q-2-R, AACCCTAACCACATGAGCAACG; Xq-F, AGCAAGCGGGTCCTGTAGTG; Xq-R, GGTGGAACTTCAGTAATCCGAAA; Xp-F, AAGAACGAAGCTTCCACAGTAT; Xp-R,

GGTGGGAGCAGATTAGAGAATAAA; GAPDH-F, AGCCACATCGCTCAGACAC; GAPDH-R, GCCCAATACGACCAAATCC; GAPDH RT, GCCCAATACGACCAAATCC; TERRA RT, CCCTAACCCTAACCCTAACCCTAACCCTAA; ZBP1-F, AACATGCAGCTACAATTCCAGA; ZBP1-R, AGTCTCGGTTCACATCTTTTGC; IFNB1-F, ACGCCGCATTGACCATCTAT; IFNB1-R, GTCTCATTCCAGCCAGTGCT; ISG15-F, CGCAGATCACCCAGAAGATCG; ISG15-R, TTCGTCGCATTTGTCCACCA; IFI44-F, AGCCGTCAGGGATGTACTATAAC; IFI44-R, AGGGAATCATTTGGCTCTGTAGA; CCL5-F, CCAGCAGTCGTCTTTGTCAC; CCL5-R, CTCTGGGTTGGCACACACTT; CXCL10-F, GTGGCATTCAAGGAGTACCTC; CXCL10-R, TGATGGCCTTCGATTCTGGATT.

Antibodies

Antibodies used were as follows: TRF2 (Karlseder laboratory), ZBP1 (Novus Biologicals, NBP1-76854 and Cell Signaling Technology, 60968), LC3 (Cell Signaling Technology, 2775 and 3868), STING (Cell Signaling Technology, 13647), CGAS (Cell Signaling Technology, D1D3G), γH2AX (Millipore, 05-636-I), Flag (Sigma-Aldrich, F1804), GAPDH (Abnova, PAB17013), TBK1 (Cell Signaling Technology, 3504S), phosphorylated TBK1 (Cell Signaling Technology, 5483S), STAT1 (Cell Signaling Technology, 9172S), phosphorylated STAT1 (Cell Signaling Technology, 9167L), IRF3 (Cell Signaling Technology, 11904S), phosphorylated IRF3 (Cell Signaling Technology, D4947S), MAVS (Cell Signaling Technology, 24930S), TOMM20 (Abcam, ab56783), TFAM (Abcam, ab119684), dsDNA (Abcam, ab27156), dsRNA (Millipore, MABE1134), phosphorylated ATR (Abcam, ab223258), phosphorylated ATM (Cell Signaling Technology, 5883), the apoptosis antibody sampler kit (Cell Signaling Technology, 9915), the necroptosis antibody sampler kit (Cell Signaling Technology, 98110) and the pyroptosis antibody sampler kit (Cell Signaling Technology, 43811).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data are archived at the Salk Institute. RNA-seq data are available at the Gene Expression Omnibus (GEO) repository under accession code GSE218396. A list of enriched gRNAs (log_2 fold change > 2) of the CRISPR–Cas9 screen is available in the Supplementary Information. Raw, uncropped images of western blots, southern blots of terminal restriction fragments and RNA dot-blots are provided in the Supplementary Information. Source data are provided with this paper.

References

1. Nassour, J. et al. Autophagic cell death restricts chromosomal instability during replicative crisis. *Nature* **565**, 659–663 (2019).

2. Hayashi, M. T., Cesare, A. J., Rivera, T. & Karlseder, J. Cell death during crisis is mediated by mitotic telomere deprotection. *Nature* **522**, 492–496 (2015).

3. Wright, W. E. & Shay, J. W. The two-stage mechanism controlling cellular senescence and immortalization. *Exp. Gerontol.* **27**, 383–389 (1992).

4. Fagagna, F. et al. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).

5. Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).

6. Davoli, T. & de Lange, T. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse

cells. *Cancer Cell* **21**, 765–776 (2012).

7. Wright, W. E., Pereira-Smith, O. M. & Shay, J. W. Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Mol. Cell. Biol.* **9**, 3088–3092 (1989).

8. Shay, J. W., Pereira-Smith, O. M. & Wright, W. E. A role for both RB and p53 in the regulation of human cellular senescence. *Exp. Cell. Res.* **196**, 33–39 (1991).

9. Lechner, M. S. et al. Human papillomavirus E6 proteins bind p53 in vivo and abrogate p53-mediated repression of transcription. *EMBO J.* **11**, 3045–3052 (1992).

10. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129–1136 (1990).

11. Lane, D. P. & Crawford, L. V. T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278**, 261–263 (1979).

12. Fu, Y. et al. Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display. *Gene* **240**, 157–163 (1999).

13. Kuriakose, T. & Kanneganti, T.-D. ZBP1: innate sensor regulating cell death and inflammation. *Trends Immunol.* **39**, 123–134 (2018).

14. Zhang, T. et al. Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. *Cell* **180**, 1115–1129 (2020).

15. Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U. & Rich, A. Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* **8**, 761–765 (2001).

16. Deigendesch, N., Koch-Nolte, F. & Rothenburg, S. ZBP1 subcellular localization and association with stress granules is controlled by its Z-

DNA binding domains. *Nucleic Acids Res.* **34**, 5007–5020 (2006).

17. Herbert, A. & Rich, A. The biology of left-handed Z-DNA. *J. Biol. Chem.* **271**, 11595–11598 (1996).

18. Takaoka, A. et al. DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).

19. Kuriakose, T. et al. ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci. Immunol.* **1**, aag2045 (2016).

20. Thapa, R. J. et al. DAI senses influenza A virus genomic RNA and activates RIPK3-dependent cell death. *Cell Host Microbe* **20**, 674–681 (2016).

21. Sridharan, H. et al. Murine cytomegalovirus IE3-dependent transcription is required for DAI/ZBP1-mediated necroptosis. *EMBO Rep.* **18**, 1429–1441 (2017).

22. Kesavardhana, S. et al. ZBP1/DAI ubiquitination and sensing of influenza vRNPs activate programmed cell death. *J. Exp. Med.* **214**, 2217–2229 (2017).

23. Maelfait, J. et al. Sensing of viral and endogenous RNA by ZBP1/DAI induces necroptosis. *EMBO J.* **36**, 2529–2543 (2017).

24. Jiao, H. et al. Z-nucleic-acid sensing triggers ZBP1-dependent necroptosis and inflammation. *Nature* **580**, 391–395 (2020).

25. Wang, R. et al. Gut stem cell necroptosis by genome instability triggers bowel inflammation. *Nature* **580**, 386–390 (2020).

26. Kesavardhana, S. et al. The Zα2 domain of ZBP1 is a molecular switch regulating influenza-induced PANoptosis and perinatal lethality during development. *J. Biol. Chem.* **295**, 8325–8330 (2020).

27. Rothenburg, S., Schwartz, T., Koch-Nolte, F. & Haag, F. Complex regulation of the human gene for the Z-DNA binding protein DLM-1. *Nucleic Acids Res.* **30**, 993–1000 (2002).

28. Ponnusamy, K. et al. The innate sensor ZBP1-IRF3 axis regulates cell proliferation in multiple myeloma. *Haematologica* **107**, 721–732 (2021).

29. Ha, S. C. et al. The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proc. Natl Acad. Sci. USA* **105**, 20671–20676 (2008).

30. Kaiser, W. J., Upton, J. W. & Mocarski, E. S. Receptor-interacting protein homotypic interaction motif-dependent control of NF-κB activation via the DNA-dependent activator of IFN regulatory factors. *J. Immunol.* **181**, 6427–6434 (2008).

31. Rebsamen, M. et al. DAI/ZBP1 recruits RIP1 and RIP3 through RIP homotypic interaction motifs to activate NF-κB. *EMBO Rep.* **10**, 916–922 (2009).

32. Xu, L.-G. et al. VISA is an adapter protein required for virus-triggered IFN-β signaling. *Mol. Cell* **19**, 727–740 (2005).

33. Meylan, E. et al. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).

34. Kawai, T. et al. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat. Immunol.* **6**, 981–988 (2005).

35. Seth, R. B., Sun, L., Ea, C.-K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669–682 (2005).

36. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801 (2007).

37. Nergadze, S. G. et al. CpG-island promoters drive transcription of human telomeres. *RNA* **15**, 2186–2194 (2009).

38. Porro, A. et al. Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat. Commun.* **5**, 5379 (2014).

39. Jiang, M. et al. Self-recognition of an inducible host lncRNA by RIG-I feedback restricts innate immune response. *Cell* **173**, 906–919 (2018).

40. Xie, Q. et al. Long noncoding RNA ITPRIP-1 positively regulates the innate immune response through promotion of oligomerization and activation of MDA5. *J. Virol.* **92**, e00507-18 (2018).

41. Wang, Z. & Lieberman, P. M. The crosstalk of telomere dysfunction and inflammation through cell-free TERRA containing exosomes. *RNA Biol.* **13**, 690–695 (2016).

42. Wang, Z. et al. Telomeric repeat-containing RNA (TERRA) constitutes a nucleoprotein component of extracellular inflammatory exosomes. *Proc. Natl Acad. Sci. USA* **112**, E6293–E6300 (2015).

43. Peisley, A., Wu, B., Yao, H., Walz, T. & Hur, S. RIG-I forms signaling-competent filaments in an ATP-dependent, ubiquitin-independent manner. *Mol. Cell* **51**, 573–583 (2013).

44. Peisley, A. et al. Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition. *Proc. Natl Acad. Sci. USA* **108**, 21010–21015 (2011).

45. Sohn, J. & Hur, S. Filament assemblies in foreign nucleic acid sensors. *Curr. Opin. Struct. Biol.* **37**, 134–144 (2016).

46. Nelson, I., Hanna, M. G., Wood, N. W. & Harding, A. E. Depletion of mitochondrial DNA by ddC in untransformed human cell lines. *Somat.*

Cell Mol. Genet. **23**, 287–290 (1997).

47. Borowski, L. S., Dziembowski, A., Hejnowicz, M. S., Stepien, P. P. & Szczesny, R. J. Human mitochondrial RNA decay mediated by PNPase-hSuv3 complex takes place in distinct foci. *Nucleic Acids Res.* **41**, 1223–1240 (2013).

48. Myong, S. et al. Cytosolic viral sensor RIG-I is a 5′-triphosphate dependent translocase on double stranded RNA. *Science* **323**, 1070–1074 (2009).

49. Berke, I. C. & Modis, Y. MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA. *EMBO J.* **31**, 1714–1726 (2012).

50. Peisley, A. et al. Kinetic mechanism for viral dsRNA length discrimination by MDA5 filaments. *Proc. Natl Acad. Sci. USA* **109**, E3340–E3349 (2012).

51. Hubel, P. et al. A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape. *Nat. Immunol.* **20**, 493–502 (2019).

52. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the Rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).

53. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

54. Campeau, E. et al. A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS ONE* **4**, e6529 (2009).

55. Zafra, M. P. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).

56. Arnoult, N. et al. Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature* **549**, 548–552 (2017).

57. Beishline, K. et al. CTCF driven TERRA transcription facilitates completion of telomere DNA replication. *Nat. Commun.* **8**, 2114 (2017).

58. Dixit, E. et al. Peroxisomes are signaling platforms for antiviral innate immunity. *Cell* **141**, 668–681 (2010).

59. O'Sullivan, R. J., Kubicek, S., Schreiber, S. L. & Karlseder, J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat. Struct. Mol. Biol.* **17**, 1218–1225 (2010).

60. Cesare, A. J. et al. Spontaneous occurrence of telomeric DNA damage response in the absence of chromosome fusions. *Nat. Struct. Mol. Biol.* **16**, 1244–1251 (2009).

61. Hendrickson, D. G., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17**, 28 (2016).

62. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

63. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

64. Feretzaki, M. & Lingner, J. A practical qPCR approach to detect TERRA, the elusive telomeric repeat-containing RNA. *Methods* **114**, 39–45 (2017).

65. O'Sullivan, R. J. et al. Rapid induction of alternative lengthening of telomeres by depletion of the histone chaperone ASF1. *Nat. Struct. Mol. Biol.* **21**, 167–174 (2014).

66. Lai, T.-P., Wright, W. E. & Shay, J. W. Generation of digoxigenin-incorporated probes to enhance DNA detection sensitivity. *BioTechniques* **60**, 306–309 (2016).

67. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

68. *FastQC: a Quality Control Tool for High Throughput Sequence Data* https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Babraham Bioinformatics, 2019).

69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

We thank N. Arnoult for advice; B. Silva and C. M. Azzalin for discussions; P. Lieberman for reagents; R. Evans for the use of their IncuCyte S3 system; S. P. Moreno for sharing protocols for necroptosis activation; and N. Luhtala for sharing THP-1 cell line for pyroptosis activation. J.N. is supported by the European Molecular Biology Organization (ALTF 213-216) and the Hewitt Foundation and the National Cancer Institute (K99CA252447). A.C., L.M. and N.T. have received support from the Glenn Foundation for Biology of Aging Research. T.T.S. has received support from the European Molecular Biology organization (ALTF 668-2019). S.P. has received support from Swiss National Science Foundation Early Postdoc Mobility Fellowship (P2ZHP3_195173). M.N.S and A.W. are supported by the NCI (P30-014194) and NIA (AG073084, AG064049 and AG068635). J.K. receives support from The Salk Institute Cancer Centre Core Grant (P30CA014195), the National Institutes of Health (RO1CA227934, RO1CA234047, RO1CA228211 and RO1AG077324), the Donald and Darlene Shiley Chair, the Samuel Waxman Cancer Research Foundation, the Highland Street Foundation and the American Heart Association (19PABHI34610000). G.S.S. receives support from the National Institutes of Health (RO1AG077324 and R01AR069876), the

American Heart Association-Allen Institute (19PABHI34610000) and the Audrey Geisel Chair in Biomedical Science.

Author information

Authors and Affiliations

1. The Salk Institute for Biological Studies, La Jolla, CA, USA

   Joe Nassour, Lucia Gutierrez Aguiar, Adriana Correia, Tobias T. Schmidt, Laura Mainz, Sara Przetocka, Candy Haggblom, Nimesha Tadepalle, April Williams, Maxim N. Shokhirev, Gerald S. Shadel & Jan Karlseder

2. Departamento de Biologia Vegetal, Faculdade de Ciências da Universidade de Lisboa (FCUL), Lisbon, Portugal

   Adriana Correia

3. A*STAR Division of Cancer Genetics, Institute of Molecular and Cell Biology (IMCB), Singapore, Singapore

   Semih C. Akincilar & Vinay Tergaonkar

4. Therapeutics Laboratory of NFκB Signaling, Institute of Molecular and Cell Biology (IMCB), Singapore, Singapore

   Semih C. Akincilar & Vinay Tergaonkar

5. Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore, Singapore

   Vinay Tergaonkar

6. Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore, Singapore

   Vinay Tergaonkar

Contributions

Experiments were designed by J.N., G.S.S. and J.K. Experiments were performed by J.N. (all experiments except for the ones outlined below), L.G.A. (those shown in Fig. 3a,b and Extended Data Figs. 7c, 8a,b and 10b,d,f), A.C. (those shown in Figs. 1d, 2b and 3c and Extended Data Figs. 4d and 12d–f), T.T.S. (those shown in Figs. 1c and 3f and Extended Data Fig. 3d, and design of sgZBP1 and ZBP1 expression vectors), L.M. (those shown in Extended Data Fig. 7b), S.P. (those shown in Extended Data Figs. 6e, 12b and 13b), S.C.A. (those shown in Extended Data Fig. 3g), and M.N.S. and A.W. (bioinformatics analysis of fRIP–seq and RNA-seq). N.T. provided experimental advice for Extended Data Fig. 11d. V.T. provided advice for Extended Data Fig. 3g. C.H. coordinated all supplies. J.N., G.S.S. and J.K. wrote the manuscript.

Corresponding authors

Correspondence to Gerald S. Shadel or Jan Karlseder.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Andreas Wack and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Crisis cells exhibit an inflammatory gene signature.

a, Heat maps representing RNA-sequencing data for DEGs between growing and crisis IMR90^{E6E7} and WI38^{SV40LT} at the indicated PDs. DEGs with a fold change > 1.5 and false-discovery rate <0.05 are shown. One experiment was performed. **b**, Venn diagram of commonly upregulated DEGs (n = 1,490) in crisis of IMR90^{E6E7} and WI38^{SV40LT} as compared with growing cells. **c**, Graphical representation of the enrichment of gene ontology terms for commonly upregulated DEGs (n = 1,490) in crisis. The top 20 biological processes are shown. Abbreviations: DEGs: differentially expressed genes; GO: gene ontology; PDs: population doublings.

Source data

Extended Data Fig. 2 ISGs are induced during crisis.

a, Venn diagram of RNA-sequencing data showing significantly upregulated ISGs in crisis of IMR90^{E6E7}, WI38^{SV40LT}, or both types of fibroblasts as compared with growing cells (PDs are indicated in Extended Data Fig. 1b). Upregulated ISGs with a fold change > 1.5 and false-discovery rate < 0.05 are shown. The complete list of ISGs includes 390 genes and is based on the RNAseq analysis of 5 cell types transfected with 20 viruses from 9 virus classes[51]. One experiment was performed. **b**, Heat maps representing RNA-sequencing data of upregulated ISGs (n = 61) in crisis of both types of fibroblasts as in **a** (lower panel). **c**, Graphical representation of the enrichment of gene ontology terms for the genome-wide CRISPR/Cas9 knock-out screen. sgRNAs with $\log_2$ fold change > 2 are used. The top 20 biological processes are shown. Two independent experiments were performed. Abbreviations: DEGs: differentially

expressed genes; sgRNA: single guide RNA; GO: gene ontology; ISGs: interferon stimulated genes, PD: population doubling.

Source data

Extended Data Fig. 3 Loss of ZBP1 attenuates IFN response and promotes crisis bypass.

a, Line chart showing growth curves analysis of IMR90^{E6E7} or WI38^{SV40LT} that have bypassed senescence expressing control sgRNAs (sg*LUC*, sg*GFP*) or three individual sgRNAs targeting *ZBP1* (sg*ZBP1*#1, sg*ZBP1*#2, sg*ZBP1*#3). Three independent experiments were performed. **b**, Crystal violet viability assay of IMR90^{E6E7} or WI38^{SV40LT} as in **a**. Cells were seeded at day 30 and fixed at day 40. One experiment was performed. **c**, Immunoblotting of IMR90^{E6E7} or WI38^{SV40LT} as in **a**. Protein extracts were collected at day 30 when control cells (sg*LUC*, sg*GFP*) entered crisis and upregulated *ZBP1*. GAPDH loading control. Two independent experiments were performed. **d**, Line chart showing IncuCyte analysis of cell death of IMR90^{E6E7} or WI38^{SV40LT} as in **a**, measured in real time by Cytotox green incorporation at day 30 of the experiment. Bars represent mean ± s.e.m. from technical replicate. One-way ANOVA, *** $p < 0.001$. Two independent experiments were performed. **e**, Scatter plots with bars showing RT-qPCR analysis of *IFNB1* in IMR90^{E6E7} or WI38^{SV40LT} as in **a**. RNA extracts were collected at day 30 of the experiment. Expression levels were normalized to control cells with sg*LUC*. Bars represent mean ± s.d. of technical replicates. n: number of technical replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. Two independent experiments were performed. **f**, Scatter plots with bars showing ELISA assays of IFN-β levels secreted by IMR90^{E6E7} or WI38^{SV40LT} as in **a**. Culture supernatant were collected at day 30 of the experiment. Bars represent mean ± s.d. of technical replicates. n: number of technical replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. One experiment was performed. Abbreviations: GFP: Green fluorescent protein; LUC: luciferase; PDs: population doublings.

Source data

[Extended Data Fig. 4 ZBP1(S) levels increase during crisis dependent on cGAS, STING and MAVS.]

a, Graphical representation of human ZBP1 protein showing the positions of Zα1, Zα2 and RHIM domains with reference to the two main ZBP1 isoforms, ZBP1(L) and ZBP1(S). **b**, Scatter plots with bars showing RT-qPCR analysis of ZBP1 in IMR90^{E6E7} and WI38^{SV40LT} at the indicated PDs. Expression levels were normalized to growing cells. Bars represent mean ± s.d. of technical replicates. n: number of technical replicates. Two-tailed Student's *t*-test, *** $p < 0.001$. Three independent experiments were performed. **c**, Immunoblotting of IMR90^{E6E7} and WI38^{SV40LT} at the indicated PDs. GAPDH loading control. Three independent experiments were performed. **d**, Immunoblotting of crisis (PD108) IMR90^{E6E7} transduced with two individual lentiviral shRNA vectors targeting either ZBP1(L) by binding to a sequence within the Zα1 or a sequence present in both ZBP1 isoforms. Protein extracts were collected at day 6 post-shRNA transfection. GAPDH loading control. Two independent experiments were performed. **e**, Scatter plots showing RT-qPCR analysis of ZBP1 in crisis (PD106) IMR90^{E6E7} transfected with siRNA targeting IFN receptors and main nucleic-acid sensors and adapters. RNA extracts were collected at day 4 post-siRNA transfection. Expression levels were normalized to siCTR cells. Bars represent mean ± s.d. from technical replicates. n: number of technical replicates. One-way ANOVA, ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Three independent experiments were performed. Abbreviations: ZBP1(L): long isoform; ZBP1(S): short isoform; CTR: control; PDs: population doublings.

[Source data]

[Extended Data Fig. 5 Induction of ISGs requires both Zα2 and RHIM1 domains of ZBP1.]

a, Schematic representation of the human ZBP1 mutants. Asterisk represents amino-acid mutation localized in Zα2, RHIM1, or RHIM2. Numbers indicate amino-acid positions. Names of the mutants are listed on the right. **b**, Immunoblotting of parental and ZBP1 CRISPR KO clones of

IMR90^{E6E7}. Protein extracts were collected 72 h after treatment with 500 U ml^{-1} of IFN-β. GAPDH loading control. Two independent experiments were performed. **c**, Immunoblotting of ZBP1 KO MR90^{E6E7} (clone 38) reconstituted with either WT or mutant forms of ZBP1(S)-Flag listed in **a**. Protein extracts were collected 14 days after transduction. GAPDH loading control. Two independent experiments were performed. **d**, Scatter plots showing RT-qPCR analysis of ISGs in ZBP1 KO IMR90^{E6E7} (clone 38) reconstituted with either WT ZBP1(S)-Flag or mutants ZBP1(S)-Flag lacking Zα2, Zα2-RHIM1-RHIM2, C-terminal region, RHIM1, RHIM2, or RHIM1-RHIM2. Expression levels were normalized to control cells with an empty vector. Bars represent mean ± s.d. of technical replicates n: number of technical replicates. One-way ANOVA, ns: not significant, * $p < 0.05$, *** $p < 0.001$. Three independent experiments were performed. Abbreviations: C: clone; P: parental; KO: knock-out; ZBP1(L): long isoform; ZBP1(S): short isoform; SE: short exposure; LE: long exposure; WT: wild-type; PD: population doubling.

Source data

Extended Data Fig. 6 ZBP1(S)-mediated cell death does not involve PANoptosis.

a, Immunoblotting of pre-crisis (PD97) IMR90^{E6E7} expressing empty vector or ZBP1(S)-Flag at day 10 post-transduction. Pre-crisis IMR90^{E6E7} treated with staurosporine (1 μM) serve as positive controls. GAPDH loading control. Two independent experiments were performed. **b**, Immunoblotting of pre-crisis (PD97) IMR90^{E6E7} expressing empty vector or ZBP1(S)-Flag at day 10 post-transduction. HT-29 cells treated with TBZ (TNF (1000 U ml^{-1}) + BV6 (1 μM) + z-VAD-FMK (20 μM)) serve as positive controls. GAPDH loading control. Two independent experiments were performed. **c**, Immunoblotting of pre-crisis (PD97) IMR90^{E6E7} expressing empty vector or ZBP1(S)-Flag at day 10 post-transduction. THP-1 cells primed with LPS (1 μg ml^{-1}) for 4 h prior to treatment with nigericin (10 μM) for 15 and 30 min serve as positive controls. GAPDH loading control. Two independent experiments were performed. **d**, Immunoblotting of pre-crisis (PD97) IMR90^{E6E7} expressing empty vector

or ZBP1(S)-Flag at day 10 post-transduction. GAPDH loading control. Two independent experiments were performed. **e**, Scatter plot with bars showing the percentage of dead pre-crisis (PD97) IMR90^{E6E7} expressing either empty vector or ZBP1(S)-Flag at day 10 post-transduction. ZBP1(S)-Flag expressing cells were transfected with siRNAs at day 8 post-transduction and incubated with Cytotox green for 24 h before imaging. Dead cells are labelled with Cytotox green and nuclei with Hoechst. Bars represent mean ± s.d. from biological replicates. n: number of biological replicates. One-way ANOVA, ns: not significant, p** < 0.01, *** p < 0.001. Three independent experiments were performed. **f**, Left: Representative fluorescence microscopy images of Metaphase-TIF assays stained with DAPI (blue), telomere FISH (green) and γH2AX immunofluorescence (red) performed on growing (PD35) IMR90^{E6E7} expressing non-targeting control shRNA or shRNA against TRF2. Mock represents non-transduced cells. Experiments were performed at the indicated days post-shRNA transduction. Scale bar 10 µm. Three independent experiments were performed. Right: Representative fluorescence microscopy images of cytogenetic preparations stained with DAPI (red) and telomere FISH (green) performed on growing (PD35) IMR90^{E6E7} expressing non-targeting control shRNA or shRNA against TRF2. Mock represents non-transduced cells. Experiments were performed at the indicated days post-shRNA transduction. Scale bar 10 µm. Three independent experiments were performed. **g**, Box and whisker plots showing the number of telomeric γH2AX foci / metaphase (left) and number of fused chromosomes / metaphase (right) at the indicated days post-shRNA transduction. Centre line: median; box limits: 1st and 3rd quartiles; whiskers: 10th and 90th percentiles. n: number of metaphases analysed. One-way ANOVA, ns: not significant, *** p < 0.001. Three independent experiments were performed. **i**, Experimental timeline of Fig. 2a. Growing (PD35) IMR90^{6E7} were first transduced with non-targeting control shRNA or shRNA against TRF2, selected, and transfected with either of two individuals siZBP1 or non-targeting control siRNA at day 4 post-shRNA transduction. Protein extracts and metaphase spreads were prepared at days 6,9, and 12 post-shRNA transduction. **j**, Box and whisker plots showing the number of fused chromosomes / metaphase at day 6 post-shRNA transduction. Centre line: median; box limits: 1st and 3rd quartiles; whiskers: 10th and 90th percentiles. n: number of metaphases analysed. One-way ANOVA, ns: not

significant. Three independent experiments were performed. Abbreviations: KO: knock-out; stauro: staurosporine; TBZ: TNF-BV6-z-VAD-FMK; LPS: lipopolysaccharide; Cl.: cleaved; Casp: caspase; NT: non-treated; TIF: telomere-dysfunction-induced foci; FISH: fluorescence in situ hybridization; CTR: control; PDs: population doublings.

Source data

Extended Data Fig. 7 Telomere dysfunction is required for ZBP1-dependent IFN signalling.

a, Line chart showing growth curves analysis of IMR90 or WI38 expressing either empty vector (pLXSN3), HPV-E6E7 or SV40-LT in combination with an empty vector (pBABE), or hTERT. Cells with empty vector (pLXSN3) are indicated as CTR. One experiment was performed. **b**, TRF assays performed on IMR90 and WI38 expressing either HPV-E6E7 or SV40-LT in combination with an empty vector (pBABE) or hTERT. Two experiments were performed. DNA extracts were collected from cells at the indicated PDs. **c**, Scatter plots showing RT-qPCR analysis of ZBP1 levels in cells as in **a**. PDs are indicated. Expression levels were normalized to control cells with the earliest PDs. Bars represent mean ± s.d. from biological replicates. n: number of biological replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. Three independent experiments were performed. Abbreviations: hTERT: human telomerase reverse transcriptase; CTR: control; PDs: population doublings.

Source data

Extended Data Fig. 8 TERRA levels increase upon telomere deprotection.

a, Left: RNA-dot blot performed on RNA isolated from total cell lysates of growing (PD40) and pre-crisis (PD96) IMR90^{E6E7} using ^{32}P-dCTP-labelled probes targeting either TERRA or GAPDH transcripts. RNase A treatment was used to assess possible DNA contamination. GAPDH loading control. Three independent experiments were performed. Right: Scatter plots with bars showing relative TERRA signal normalized to GAPDH and growing

cells. Bars represent mean ± s.d. of biological replicates. n: number of biological replicates. Two-tailed Student's *t*-test, *** $p < 0.001$. Three independent experiments were performed. **b**, Left: RNA-dot blot performed on RNA isolated from total cell lysates of growing (PD40) IMR90^{E6E7} expressing non-targeting control shRNA or shRNA against TRF2 using ^{32}P-dCTP-labelled probes targeting either TERRA or GAPDH transcripts. GAPDH loading control. RNA extracts were collected at day 6 post-shRNA transduction. RNase A treatment was used to assess possible DNA contamination. GAPDH loading control. Three independent experiments were performed. Right: Scatter plots with bars showing relative TERRA signals normalized to GAPDH and sh-scramble cells. Bars represent mean ± s.d. of biological replicates. n: number of biological replicates. Two-tailed Student's t-test, *** $p < 0.001$. Three independent experiments were performed. **c**, Left: Immunoblotting of growing (PD45) IMR90^{E6E7} expressing WT ZBP1(S)-Flag and transduced with non-targeting control shRNA or shRNA against TRF2. Total protein extracts were collected at day 12 post-shRNA transduction. GAPDH loading control. Two independent experiments were performed. Right: Immunoblotting growing (PD45) IMR90^{E6E7} as in left. Protein extracts were immunoprecipitated with anti-Flag magnetic beads at day 12 post-shRNA transduction. The Input represents total protein extracts, and the Flow-Through represents unbound fractions. **d**, Column bar graph showing the average ratio fRIP / input ($\log_2$) at individual subtelomeres. fRIP and input samples were normalized into the same sequencing depth. Bars represent mean. Two independent experiments were performed. **e**, Left: Immunoblotting of growing (PD40) IMR90^{E6E7} expressing either WT ZBP1(S)-Flag or mutant ZBP1(S)-Flag lacking Zα2 transduced with non-targeting control shRNA or shRNA against TRF2. Total protein extracts were collected at day 12 post-shRNA transduction. GAPDH loading control. Three independent experiments were performed. Right: Immunoblotting of growing (PD40) IMR90^{E6E7} as in left. Protein extracts were immunoprecipitated with anti-Flag magnetic beads at day 12 post-shRNA transduction. The Input represents total protein extracts, and the Flow-Through represents unbound fractions. Three independent experiments were performed. Abbreviations: ZBP1(S): short isoform; Chr: chromosome; WT: wild-type; PDs: population doublings.

Source data

Extended Data Fig. 9 Sensing of TERRA by ZBP1 triggers an IFN response.

a, Immunoblotting of growing (PD45) IMR90^{E6E7} expressing either WT ZBP1(S)-Flag or mutant ZBP1(S)-Flag lacking Zα2 with either TRF1(ΔN) or VP16-TRF1(ΔN). Protein extracts were collected at day 4 post-doxycycline treatment (1 µg / ml). GAPDH loading control. Two independent experiments were performed. **b**, RNA-dot blot performed on RNA isolated from cells as in **a** using ^{32}P-dCTP-labelled probes targeting either TERRA or GAPDH transcripts. RNA extracts were collected at day 4 post-doxycycline treatment (1 µg ml^{-1}). GAPDH loading control. RNase A treatment was used to assess possible DNA contamination. Two independent experiments were performed. **c**, Scatter plot with bars showing RT-qPCR analysis with 5 pairs of subtelomere-specific primers designed to measure TERRA transcribed from individual chromosome ends. Expression levels were normalized to control cells expressing TRF1(ΔN). RNA extracts were collected from cells as in **a**. Bars represent mean ± s.d. from technical replicates. n: number of technical replicates. One-way ANOVA, ** $p < 0.01$, *** $p < 0.001$. Two independent experiment were performed. Abbreviations: ZBP1(S): short isoform; Dox: doxycycline; WT: wild-type; fRIP: formaldehyde RNA immunoprecipitation; PDs: population doublings.

Source data

Extended Data Fig. 10 Sensing of TERRA by ZBP1 triggers an IFN response.

a, Line chart showing growth curve analysis of IMR90^{E6E7} expressing either WT ZBP1(S)-Flag or ZBP1(S)-Flag lacking Zα2 with either TRF1(ΔN) or VP16-TRF1(ΔN). Fresh media with doxycycline (1 µg ml^{-1}) was added every 2 days. Two independent experiments were performed. **b**, Scatter plots showing RT-qPCR analysis of ISG levels in cells as in **a**. Expression levels were normalized to control cells expressing either WT ZBP1(S)-Flag or ZBP1(S)-Flag lacking Zα2 with TRF1(ΔN). RNA extracts

were collected at day 20 of the experiment. Bars represent mean ± s.d. from technical replicates. n: number of technical replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. Two independent experiments were performed. **c**, Line chart showing growth curve analysis of IMR90^{E6E7} expressing either WT ZBP1(S)-Flag or ZBP1(S)-Flag containing point mutations in Zα2, with either TRF1(ΔN) or VP16-TRF1(ΔN). Fresh media including 1 μg ml^{-1} of doxycycline was added every 2 days. Two independent experiments were performed. **d**, Scatter plots showing RT-qPCR analysis of ISG levels in cells as in **c**. Expression levels were normalized to control cells expressing either WT ZBP1(S)-Flag or ZBP1(S)-Flag containing point mutations in Zα2 with TRF1(ΔN). RNA extracts were collected at day 30 of the experiment. Bars represent mean ± s.d. from technical replicates. n: number of technical replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. Two independent experiments were performed. **e**, Experimental timeline. Growing (PD35) IMR90^{E6E7} expressing WT ZBP1(S)-Flag were treated with doxycycline (1 μg ml^{-1}) prior to four sequential siRNA transfections. Two individual siRNAs targeting TERRA were used. Samples were collected at days 2,4,6, and 8 post-doxycycline. **f**, RNA-dot blot performed on RNA isolated from total cell lysates at day 8 post-doxycycline as shown in the timeline of the experiment, using ^{32}P-dCTP-labelled probes targeting either TERRA or GAPDH transcripts. RNase A treatment was used to assess possible DNA contamination. GAPDH loading control. Three independent experiments were performed. **g**, Western blotting performed on protein extracts collected at days 2,4,6, and 8 post-doxycycline treatment as shown in **e**. GAPDH loading control. Two independent experiments were performed. Abbreviations: ZBP1(S): short isoform; WT: wild-type; CTR: control; PDs: population doublings.

Source data

Extended Data Fig. 11 Assembly of ZBP1 filaments at mitochondria does not require sensing of mtDNA or mtRNA.

a, Upper panel: Representative confocal images of growing (PD50) IMR90^{E6E7} expressing either WT ZBP1(S)-Flag or ZBP1(S)-Flag lacking

Zα2 with either TRF1(ΔN) or VP16-TRF1(ΔN). Interphase nuclei were stained with DAPI (blue), telomere FISH (green), and γH2AX immunofluorescence (red). Scale bar 5 μm. Middle panel: Representative fluorescence microscopy images of Metaphase-TIF assays stained with DAPI (blue), telomere FISH (green) and γH2AX immunofluorescence (red). Scale bar 10 μm. Lower panel: Representative fluorescence microscopy images of cytogenetic preparations stained with DAPI (red) and telomere FISH (green). Scale bar 10 μm. Three independent experiments were performed at day 8 post-doxycycline treatment (1 μg ml^{-1}). **b**, Immunoblotting of cells as in **a** either untreated or at days 4 and 8 post-doxycycline treatment (1 μg ml^{-1}). Cells treated with 1 μM camptothecin (CPT) for 6 h serve as a positive control. GAPDH loading control. Two independent experiments were performed. **c**, Left: Representative confocal microscopy images of growing and crisis $IMR90^{E6E7}$ co-immunostained for endogenous ZBP1 and mitochondrial markers TOM20, TFAM, and MitoTracker at the indicated PDs. Scale bar 10 μm. Right: Scatter plots showing the number of ZBP1 filaments colocalizing with either TOM20, TFAM, or MitoTracker in growing and crisis $IMR90^{E6E7}$ and $WI38^{SV40LT}$ at the indicated PDs. Centre line: mean. n: number of cells analysed. One-way ANOVA, *** $p < 0.001$. Three independent experiments were performed. **d**, Left: Representative confocal microscopy images of crisis (PD100) $IMR90^{E6E7}$ immunostained with antibodies against endogenous ZBP1, dsDNA, and dsRNA. Cells were incubated with 5 μM of 2',3' dideoxycytidine (ddC) for 14 days. Non-treated cells serve as controls. Scale bar 10 μm. Three independent experiments were performed. Right: Scatter plots showing the signal intensity of dsDNA (left), the signal intensity of dsRNA (middle), and the number of ZBP1 filaments (right). Centre line: mean. n: number of cells analysed. Three independent experiments were performed. Two-tailed Student's t-test, ns: not significant, *** $p < 0.001$. Three independent experiments were performed. **e**, Left: Representative confocal microscopy images of growing (PD100) $IMR90^{E6E7}$ transfected with either non-targeting control siRNA (siCTR), siSUV3, or PNPase, and co-immunostained for endogenous ZBP1 and double-stranded RNA (dsRNA). Experiment was performed at day 7 post-siRNA transfection. Scale bar 10 μm. Three independent experiments were performed. Right: Scatter plots showing the signal intensity of dsRNA (left) and the number of ZBP1 filaments (right). Centre line: mean. n: number of

cells analysed. Two-tailed Student's t-test, ns: not significant, *** $p < 0.001$. Three independent experiments were performed. Abbreviations: ddC: dideoxycytidine; dsDNA: double-stranded DNA; dsRNA: double-stranded RNA; Mito: mitochondria; WT: wild-type; CTR: control; TIF: telomere-dysfunction-induced foci; FISH: fluorescence in situ hybridization; CPT: camptothecin; WT: wild-type; PDs: population doublings.

Source data

Extended Data Fig. 12 TERRA triggers a MAVS-dependent IFN response.

a, Experimental timeline. Growing (PD35) IMR90^{E6E7} expressing WT ZBP1(S)-Flag with either TRF1(ΔN) or VP16-TRF1(ΔN) were treated with doxycycline (1 μg ml^{-1}) prior to two sequential siRNA transfections. Samples were collected at day 8 post-doxycycline. **b**, Immunoblotting of cells as in **a** at day 8 post-doxycycline. GAPDH loading control. Three independent experiments were performed. **c**, Scatter plots showing RT-qPCR analysis of ISGs in cells as in **b**. Expression levels were normalized to control cells TRF1(ΔN) transfected with non-targeting control (siCTR) siRNA. Bars represent mean ± s.d. from technical replicates. n: number of technical replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. Three independent experiments were performed. **d**, Line chart showing growth curves analysis of IMR90^{E6E7} that have bypassed senescence expressing control sgRNAs (sg*GFP*) or sgRNAs targeting *MAVS or STING*. Two independent experiments were performed. **e**, Immunoblotting of IMR90^{E6E7} as in **a**. Protein extracts were collected at day 30 when control cells entered crisis. GAPDH loading control. Two independent experiments were performed. **f**, Heatmap representing RNA-sequencing data for ISGs in IMR90^{E6E7} as in **a**. RNA extracts were collected at day 30 when control cells entered crisis. ISGs that are significantly downregulated (sgMAVS versus sgGFP) with a fold change > 1.5 and false-discovery rate <0.05 are shown. The complete list of ISGs includes 390 genes and is based on the RNAseq analysis of 5 cell types transfected with 20 viruses from 9 virus classes[51]. One experiment was performed. Abbreviations: ZBP1(S): short

isoform; sgRNA: single guide RNA; CTR: control; WT: wild-type; PDs: population doublings.

Source data

Extended Data Fig. 13 ZBP1-MAVS axis promotes IFN-driven cell death.

a, Graphical representation of ZBP1(S) harbouring the mitochondrial targeting sequence of FIS1 with reference to the two main ZBP1 isoforms, ZBP1(L) and ZBP1(S). **b**, Immunoblotting of growing (PD66) IMR90^{E6E7} expressing either empty vector, Flag-ZBP1(S), or Flag-ZBP1(S)-FIS1. Cells were transfected with either siMAVS or non-targeting control siRNA at day 10 post-transduction and protein extracts were collected at day 14. GAPDH loading control. Two independent experiments were performed. **c**, Immunoblotting of growing (PD45) IMR90^{E6E7} (left) and post-senescence (PD23) HMECs expressing either empty vector or Flag-ZBP1(S)-FIS1. Cells were transfected with either siMAVS, siIFNAR2, or non-targeting control siRNA at day 10 post-transduction and protein extracts were collected at day 14. GAPDH loading control. Two independent experiments were performed. **d**, Scatter plot with bars showing the percentage of dead cells as in **c** after 24 h of incubation with Cytotox. Cells were transfected with siRNAs at day 8 post-transduction and incubated with Cytotox for 24 h before imaging. Dead cells are labelled with Cytotox red and nuclei with Hoechst. Bars represent mean ± s.d. from biological replicates. One-way ANOVA, ns: not significant, *** $p < 0.001$. Three independent experiments were performed. **e**, Crisis-associated cell death requires the simultaneous activation of two distinct, but functionally interconnected innate immune sensing pathways. cGAS recognizes nuclear DNA species released to the cytosol as byproducts breakage-fusion-bridge (BFB) cycles and primes the expression of several ISGs. ZBP1 becomes induced and further enhances the innate immune response by detecting TERRA molecules stemming from deprotected telomeres. This triggers ZBP1 self-oligomerizing into filaments at mitochondria capable of activating MAVS. Activation of DNA- and RNA-sensing pathways results in a chronic production of type I IFNs and other inflammatory cytokines responsible for the elimination of aberrant cells with unstable telomeres. Abbreviations: ZBP1(S): short

isoform; CTR: control; WT: wild-type; MN: micronuclei; IFNs: interferons; PDs: population doublings.

Source data

Supplementary information

Supplementary Information

Supplementary Figs. 1–6. Supplementary Figs. 1–5 contain the raw, uncropped images of the western blots, southern blots of terminal restriction fragments and RNA dot-blots. The molecular mass markers shown throughout this document are in kDa. Supplementary Fig. 6 shows a list of enriched gRNAs ($\log_2$ fold change > 2) of the CRISPR–Cas9 screen. Data represent the $\log_2$-transformed fold change of read counts before (day 0) and after enrichment (day 15).

Reporting Summary

Peer Review File

Source data

Source Data Figs. 1–4 and Source Data Extended Data Figs. 1–13.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Nassour, J., Aguiar, L.G., Correia, A. *et al.* Telomere-to-mitochondria signalling by ZBP1 mediates replicative crisis. *Nature* **614**, 767–773 (2023). https://doi.org/10.1038/s41586-023-05710-8

- Received: 30 September 2021
- Accepted: 05 January 2023
- Published: 08 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05710-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

From shortening telomeres to replicative crisis

- Paulina Strzyz

Nature Reviews Molecular Cell Biology Research Highlight 21 Feb 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05710-8

| Section menu | Main menu |

- Article
- Open Access
- Published: 22 February 2023

De novo design of luciferases using deep learning

- Andy Hsien-Wei Yeh ,
- Christoffer Norn,
- Yakov Kipnis,
- Doug Tischer,
- Samuel J. Pellock,
- Declan Evans,
- Pengchen Ma,
- Gyu Rie Lee,
- Jason Z. Zhang,
- Ivan Anishchenko,
- Brian Coventry,
- Longxing Cao,
- Justas Dauparas,
- Samer Halabiya,
- Michelle DeWitt,
- Lauren Carter,
- K. N. Houk &
- …
- David Baker

Show authors

Nature **volume 614**, pages 774–780 (2023)

- 20k Accesses

- 428 Altmetric
- Metrics details

Abstract

De novo enzyme design has sought to introduce active sites and substrate-binding pockets that are predicted to catalyse a reaction of interest into geometrically compatible native scaffolds[1,2], but has been limited by a lack of suitable protein structures and the complexity of native protein sequence–structure relationships. Here we describe a deep-learning-based 'family-wide hallucination' approach that generates large numbers of idealized protein structures containing diverse pocket shapes and designed sequences that encode them. We use these scaffolds to design artificial luciferases that selectively catalyse the oxidative chemiluminescence of the synthetic luciferin substrates diphenylterazine[3] and 2-deoxycoelenterazine. The designed active sites position an arginine guanidinium group adjacent to an anion that develops during the reaction in a binding pocket with high shape complementarity. For both luciferin substrates, we obtain designed luciferases with high selectivity; the most active of these is a small (13.9 kDa) and thermostable (with a melting temperature higher than 95 °C) enzyme that has a catalytic efficiency on diphenylterazine ($k_{cat}/K_m = 10^6$ M^{-1} s^{-1}) comparable to that of native luciferases, but a much higher substrate specificity. The creation of highly active and specific biocatalysts from scratch with broad applications in biomedicine is a key milestone for computational enzyme design, and our approach should enable generation of a wide range of luciferases and other enzymes.

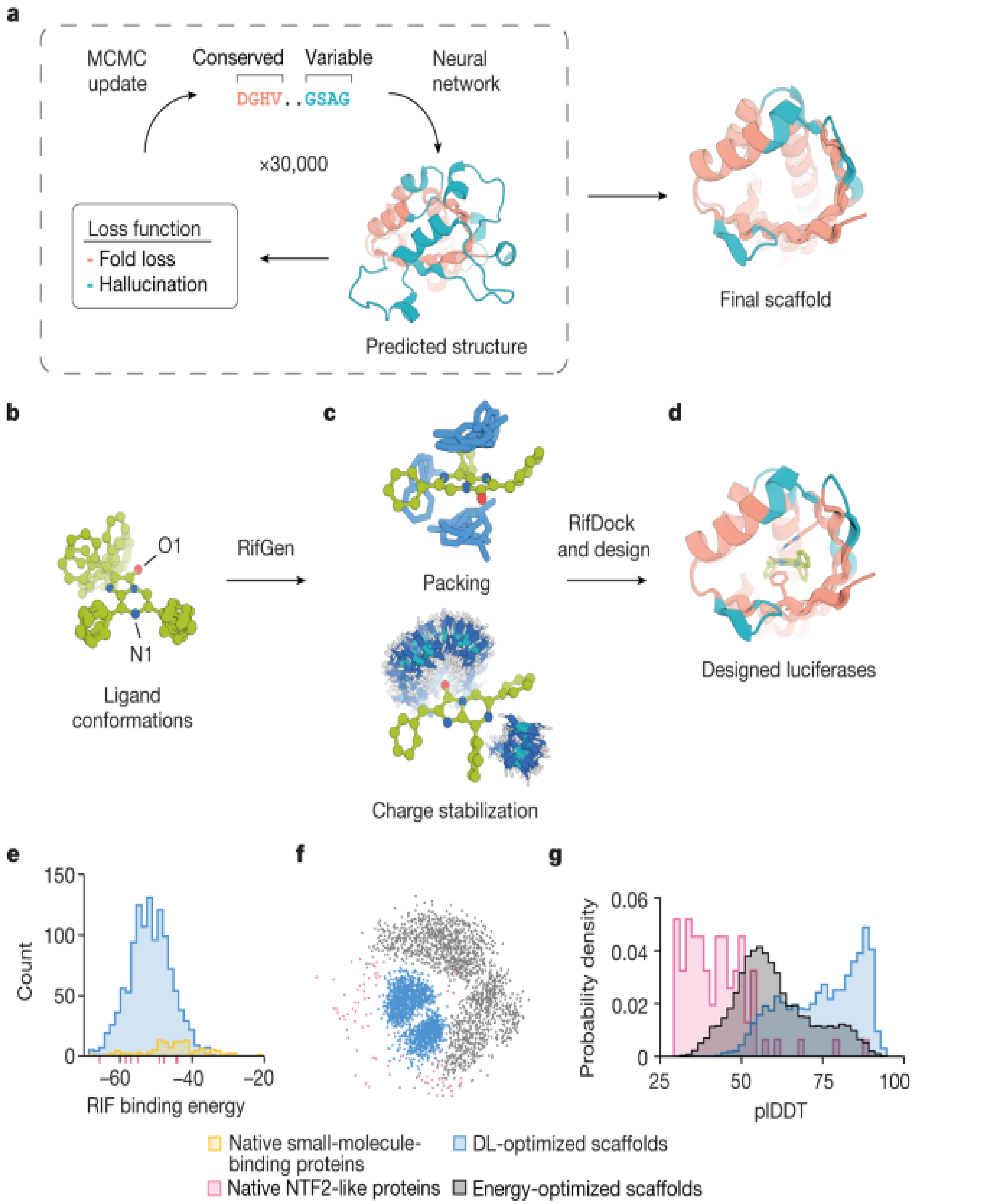
Main

Bioluminescent light produced by the enzymatic oxidation of a luciferin substrate by luciferases is widely used for bioassays and imaging in biomedical research. Because no excitation light source is needed, luminescent photons are produced in the dark; this results in higher sensitivity than fluorescence imaging in live animal models and in biological samples in which autofluorescence or phototoxicity is a concern[4,5].

However, the development of luciferases as molecular probes has lagged behind that of well-developed fluorescent protein toolkits for a number of reasons: (i) very few native luciferases have been identified[6,7]; (ii) many of those that have been identified require multiple disulfide bonds to stabilize the structure and are therefore prone to misfolding in mammalian cells[8]; (iii) most native luciferases do not recognize synthetic luciferins with more desirable photophysical properties[9]; and (iv) multiplexed imaging to follow multiple processes in parallel using mutually orthogonal luciferase–luciferin pairs has been limited by the low substrate specificity of native luciferases[10,11].

We sought to use de novo protein design to create luciferases that are small, highly stable, well-expressed in cells, specific for one substrate and need no cofactors to function. We chose a synthetic luciferin, diphenylterazine (DTZ), as the target substrate because of its high quantum yield, red-shifted emission[3], favourable in vivo pharmacokinetics[12,13] and lack of required cofactors for light emission. Previous computational enzyme design efforts have primarily repurposed native protein scaffolds in the Protein Data Bank (PDB)[1,2], but there are few native structures with binding pockets appropriate for DTZ, and the effects of sequence changes on native proteins can be unpredictable (designed helical bundles have also been used as enzyme scaffolds[14,15,16], but these are limited in number and most do not have pockets that are suitable for DTZ binding). To circumvent these limitations, we set out to generate large numbers of small and stable protein scaffolds with pockets of the appropriate size and shape for DTZ, and with clear sequence–structure relationships to facilitate subsequent active-site incorporation. To identify protein folds that are capable of hosting such pockets, we first docked DTZ into 4,000 native small-molecule-binding proteins. We found that many nuclear transport factor 2 (NTF2)-like folds have binding pockets with appropriate shape complementarity and size for DTZ placement (pink dashes in Fig. 1e), and hence selected the NTF2-like superfamily as the target topology.

Fig. 1: Generation of idealized scaffolds and computational design of de novo luciferases.



a, Family-wide hallucination. Sequences encoding proteins with the desired topology are optimized by Markov chain Monte Carlo (MCMC) sampling with a multicomponent loss function. Structurally conserved regions (peach) are evaluated on the basis of consistency with input residue–residue distance and orientation distributions obtained from 85 experimental structures of

NTF2-like proteins, whereas variable non-ideal regions (teal) are evaluated on the basis of the confidence of predicted inter-residue geometries calculated as the KL divergence between network predictions and the background distribution. The sequence-space MCMC sampling incorporates both sequence changes and insertions and deletions (see Supplementary Methods) to guide the hallucinated sequence towards encoding structures with the desired folds. Hydrogen-bonding networks are incorporated into the designed structures to increase structural specificity. **b**–**d**, The design of luciferase active sites. **b**, Generation of luciferase substrate (DTZ) conformers. **c**, Generation of a Rotamer Interaction Field (RIF) to stabilize anionic DTZ and form hydrophobic packing interactions. **d**, Docking of the RIF into the hallucinated scaffolds, and optimization of substrate–scaffold interactions using position-specific score matrices (PSSM)-biased sequence design. **e**, Selection of the NTF2 topology. The RIF was docked into 4,000 native small-molecule-binding proteins, excluding proteins that bind the luciferin substrate using more than five loop residues. Most of the top hits were from the NTF2-like protein superfamily (pink dashes). Using the family-wide hallucination scaffold generation protocol, we generated 1,615 scaffolds and found that these yielded better predicted RIF binding energies than the native proteins. **f**,**g**, Our DL-optimized scaffolds sample more within the space of the native structures (**f**) and have stronger sequence-to-structure relationships (more confident Alphafold2 structure predictions) (**g**) than native or previous non-deep-learning energy-optimized scaffolds.

Family-wide hallucination

Native NTF2 structures have a range of pocket sizes and shapes but also contain features that are not ideal, such as long loops that compromise stability. To create large numbers of ideal NTF2-like structures, we developed a deep-learning-based ‘family-wide hallucination’ approach that integrates unconstrained de novo design[17,18] and Rosetta sequence-design approaches[19] to enable the generation of an essentially unlimited number of proteins that have a desired fold (Fig. 1a). The family-wide hallucination approach used the de novo sequence and structure discovery capability of unconstrained protein hallucination[17,18] for loop and variable regions, and structure-guided sequence optimization for core regions. We used the trRosetta structure prediction neural network[20], which is effective in

identifying experimentally successful de-novo-designed proteins and hallucinating new globular proteins of diverse topologies. Starting from the sequences of 2,000 naturally occurring NTF2s, we carried out Monte Carlo searches in sequence space, at each step making a sequence change and predicting the structure using trRosetta. As the loss function guiding search, we used the confidence of the neural network in the predicted structure (as in our previous free hallucination study) supplemented with a topology-specific loss function over core residue pair geometries (see Supplementary Methods); in the loop regions, we also allowed the number of residues to vary, which resulted in short near ideal loops. To further encode structural specificity, we incorporated buried, long-range hydrogen-bonding networks. The resulting 1,615 family-wide hallucinated NTF2 scaffolds provided more shape-complementary binding pockets for DTZ than did native small-molecule-binding proteins (Fig. 1e). This method samples protein backbones that are closer to native NTF2-like proteins (Fig. 1f) and that have better scaffold quality metrics than those produced in a previous non-deep-learning energy-based approach[21] (Fig. 1g).

De novo design of luciferases for DTZ

Computational enzyme design generally starts from an ideal active site or theozyme consisting of protein functional groups surrounding the reaction transition state that is then matched into a set of existing scaffolds[1,2]. However, the detailed catalytic geometry of native marine luciferases is not well understood because only a handful of apo structures and no holo structures with luciferin substrates have been solved (at the time of this study)[22,23,24]. Both quantum chemistry calculations[25,26] and experimental data[27,28] suggest that the chemiluminescent reaction proceeds through an anionic species and that the polarity of the surroundings can substantially alter the free energy of the subsequent single-electron transfer (SET) process with triplet molecular oxygen ($^3O_2$). Guided by these data (Extended Data Fig. 1), we sought to design a shape-complementary catalytic site that stabilizes the anionic state of DTZ and lowers the SET energy barrier, assuming that the downstream dioxetane light emitter thermolysis steps are spontaneous. To stabilize the anionic state, we focused on the placement of

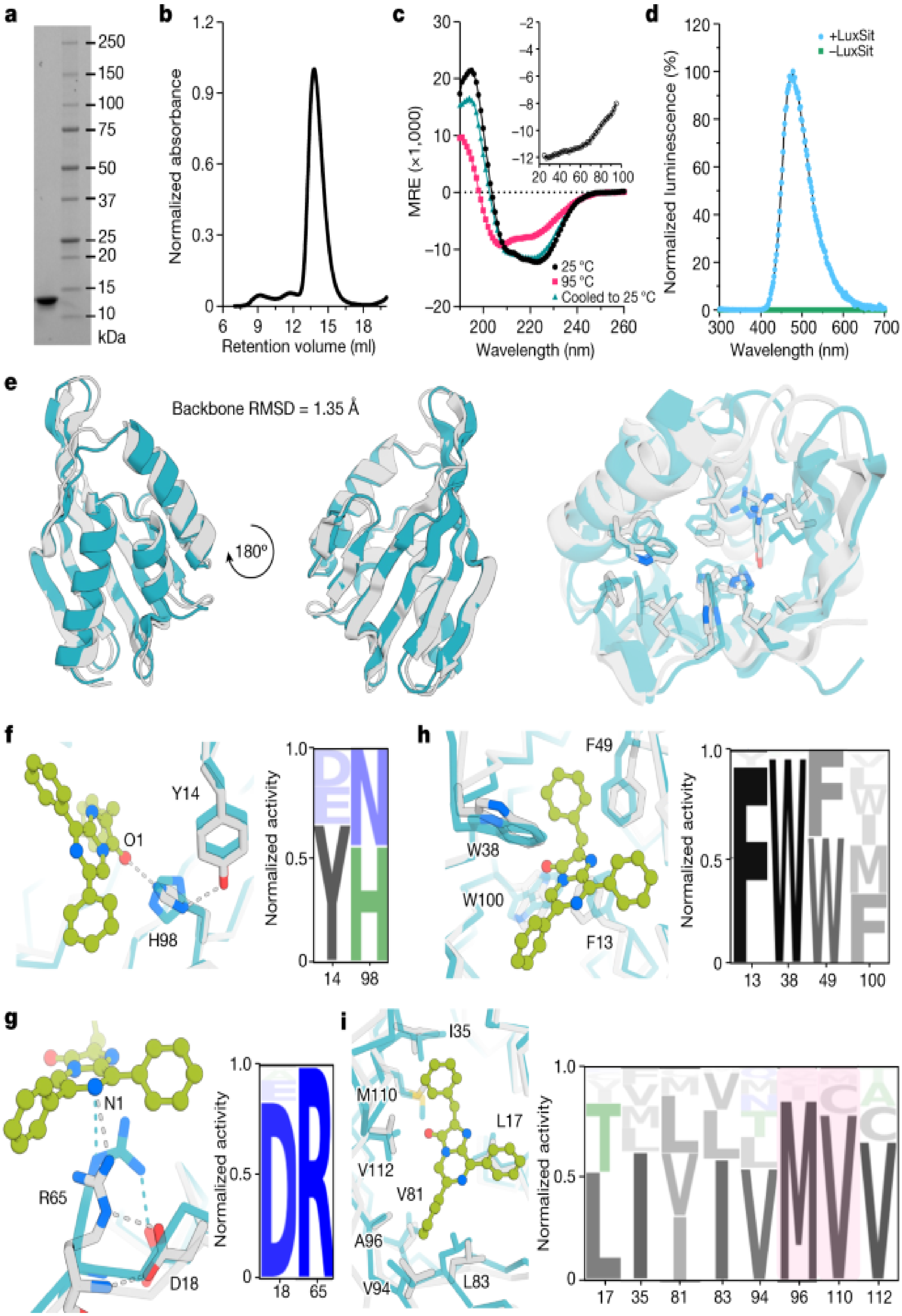
the positively charged guanidinium group of an arginine residue to stabilize the developing negative charge on the imidazopyrazinone group.

To computationally design such active sites into large numbers of hallucinated NTF2 scaffolds, we first generated an ensemble of anionic DTZ conformers (Fig. 1b). Next, around each conformer, we used the RifGen method[29,30] to enumerate rotamer interaction fields (RIFs) on three-dimensional grids consisting of millions of placements of amino acid side chains making hydrogen-bonding and nonpolar interactions with DTZ (Fig. 1c). An arginine guanidinium group was placed adjacent to the N1 atom of the imidazopyrazinone group to stabilize the negative charge. RifDock was then used to dock each DTZ conformer and associated RIF in the central cavity of each scaffold to maximize protein–DTZ interactions. An average of eight side-chain rotamers, including an arginine residue to stabilize the anionic imidazopyrazinone core, were positioned in each pocket (Supplementary Fig. 2a). For the top 50,000 docks with the most favourable side chain–DTZ interactions, we optimized the remainder of the sequence using RosettaDesign (Fig. 1d) for high-affinity binding to DTZ with a bias towards the naturally observed sequence variation to ensure foldability. During the design process, pre-defined hydrogen-bond networks (HBNets) in the scaffolds were kept intact for structural specificity and stability, and interactions of these HBNet side chains with DTZ were explicitly required in the RifDock step to ensure the preorganization of residues that are essential for catalysis. In the first sequence-design step, the identities of all RIF and HBNet residues were kept fixed, and the surrounding residues were optimized to hold the side chain–DTZ interactions in place and maintain structural specificity. In the second sequence-design step, the RIF residue identities (except the arginine) were also allowed to vary, as Rosetta can identify apolar and aromatic packing interactions that were missed in the RIF owing to binning effects. During sequence design, the scaffold backbone, side chains and DTZ substrate were allowed to relax in Cartesian space. After sequence optimization, the designs were filtered on the basis of ligand-binding energy, protein–ligand hydrogen bonds, shape complementarity and contact molecular surface, and 7,648 designs were selected and ordered as pooled oligos for experimental screening.

Identification of active luciferases

Oligonucleotides encoding the two halves of each design were assembled into full-length genes and cloned into an *Escherichia coli* expression vector (see Supplementary Methods). A colony-based screening method was used to directly image active luciferase colonies from the library and the activities of selected clones were confirmed using a 96-well plate expression (Extended Data Fig. 2). Three active designs were identified; we refer to the most active of these as LuxSit (from the Latin lux sit, 'let light exist'), which at 117 residues (13.9 kDa) is, to our knowledge, smaller than any previously described luciferase. Biochemical analysis, including SDS–PAGE and size-exclusion chromatography (Fig. 2a,b and Extended Data Fig. 3), indicated that LuxSit is highly expressed in *E. coli*, soluble and monomeric. Circular dichroism (CD) spectroscopy showed a strong far-ultraviolet CD signature, suggesting an organized α-β structure. CD melting experiments showed that the protein is not fully unfolded at 95 °C, and that the full structure is regained when the temperature is dropped (Fig. 2c). Incubation of LuxSit with DTZ resulted in luminescence with an emission peak at around 480 nm (Fig. 2d), consistent with the DTZ chemiluminescence spectrum. Although we were not able to determine the crystal structure of LuxSit, the structure predicted by AlphaFold2 (ref. [31]) is very close to the design model at the backbone level (root-mean-square deviation (RMSD) = 1.35 Å) and over the side chains interacting with the substrate (Fig. 2e). The designed LuxSit active site contains Tyr14–His98 and Asp18–Arg65 dyads, with the imidazole nitrogen atoms of His98 making hydrogen-bond interactions with Tyr14 and the O1 atom of DTZ (Fig. 2f). The centre of the Arg65 guanidinium cation is 4.2 Å from the N1 atom of DTZ and Asp18 forms a bidentate hydrogen bond to the guanidinium group and backbone N–H of Arg65 (Fig. 2g).

Fig. 2: Biophysical characterization of LuxSit.



a
250
150
100
75
50
37
25
20
15
10
kDa
b
Normalized absorbance
Retention volume (ml)
c
MRE (×1,000)
Wavelength (nm)
25 °C
95 °C
Cooled to 25 °C
d
+LuxSit
–LuxSit
Normalized luminescence (%)
Wavelength (nm)
e
Backbone RMSD = 1.35 Å
180°
f
Y14
O1
H98
Normalized activity
h
F49
W38
W100
F13
Normalized activity
g
N1
R65
D18
Normalized activity
i
I35
M110
L17
V112
V81
A96
V94
L83
Normalized activity

a, Coomassie-stained SDS–PAGE of purified recombinant LuxSit from *E. coli* (for gel source data, see Supplementary Fig. 1). **b**, Size-exclusion chromatography of purified LuxSit suggests monodispersed and monomeric properties. **c**, Far-ultraviolet CD spectra at 25 °C (black), 95 °C (red) and cooled back to 25 °C (green). Insert, CD melting curve of LuxSit at 220 nm. MRE, molar residue ellipticity. **d**, Luminescence emission spectra of DTZ in the presence (blue) and absence (green) of LuxSit. **e**, Structural alignment of the design model (blue) and AlphaFold2-predicted model (grey), which are in close agreement at both the backbone (left) and the side-chain (right) level. **f**–**i**, Site-saturation mutagenesis of substrate-interacting residues. Magnified views (left) of designed (blue) and AlphaFold2 (grey) models at the side-chain level, illustrating the designed enzyme–substrate interactions of Tyr14–His98 core HBNet (**f**), Asp18–Arg65 dyad (**g**), π-stacking (**h**) and hydrophobic packing (**i**) residues. Sequence profiles (right) are scaled by the activities of different sequence variants: (activity for the indicated amino acid)/(sum of activities over all tested amino acids at the indicated position). A96M and M110V substitutions with increased activity are highlighted in pink.

Source data

De novo design of luciferases for h-CTZ

We next sought to apply the knowledge gained from designing LuxSit to create 2-deoxycoelenterazine (h-CTZ)-specific luciferases. Because the molecular shape of h-CTZ is different from that of DTZ, we created an additional set of NTF2 superfamily scaffolds (see Supplementary Methods) with matching pocket shapes and high model confidence (AlphaFold2-predicted local-distance difference test (pLDDT) > 92). We then installed catalytic sites in these scaffolds and designed the first shell-protein side chain–h-CTZ interactions using the histidine and arginine substrate interaction geometries that were most successful in the first round for DTZ. To design the remainder of the sequence, we used ProteinMPNN[32], which can result in better stability, solubility and accuracy than RosettaDesign. After filtering on the basis of the AlphaFold2-predicted pLDDT, Cα RMSD, contact molecular surface and Rosetta-computed binding energies (see Supplementary Methods), we selected and experimentally expressed 46

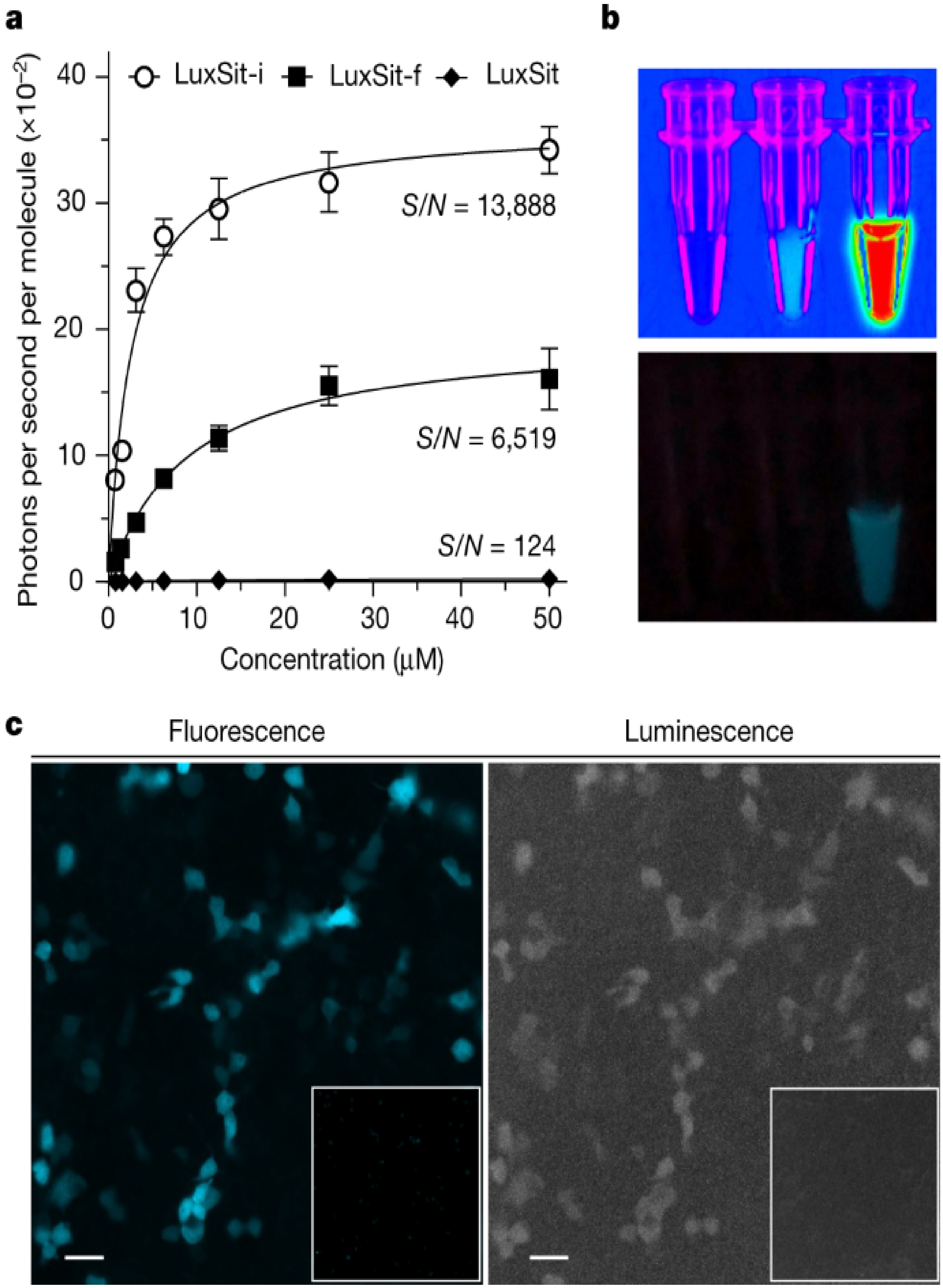
designs in *E. coli* and identified 2 (HTZ3-D2 and HTZ3-G4) that had luciferase activity with the h-CTZ luciferin substrate. Both designs were highly soluble, monodisperse and monomeric, and the luciferase activities were of the same order of magnitude as LuxSit (Extended Data Fig. 4). The success rate increased from 3/7,648 to 2/46 sequences in the second round, probably owing to the knowledge of active-site geometry from the first round and the increased robustness of the ProteinMPNN method of sequence design.

Optimization of luciferase activity

To better understand the contributions to the catalysis of LuxSit, the most active of our designs, we constructed a site-saturation mutagenesis (SSM) library in which each residue in the substrate-binding pocket was mutated to every other amino acid one at a time (see Supplementary Methods), and determined the effect of each mutation onluciferase activity. Figure 2f–i shows the amino acid preferences at key positions. Arg65 is highly conserved (Fig. 2g), and its dyad partner Asp18 can only be mutated to Glu (which reduces activity), suggesting that the carboxylate–Arg65 hydrogen bond is important for luciferase activity. In the Tyr14–His98 dyad (Fig. 2f), Tyr14 can be substituted with Asp and Glu, and His98 can be replaced with Asn. As all active variants had hydrogen-bond donors and acceptors at these positions, the dyads might help to mediate the electron and proton transfer required for luminescence. Hydrophobic (Fig. 2i) and π-stacking (Fig. 2h) residues at the binding interface tolerate other aromatic or aliphatic substitutions and generally prefer the amino acid in the original design, consistent with model-based affinity predictions of mutational effects (Extended Data Fig. 5). The A96M and M110V mutants (highlighted in pink) increase activity by 16-fold and 19-fold, respectively, over LuxSit (Supplementary Table 1). Optimization guided by these results yielded LuxSit-f (A96M/M110V), with a flash-type emission kinetic, and LuxSit-i (R60S/A96L/M110V), with a photon flux more than 100-fold higher than that of LuxSit (Extended Data Fig. 6). Overall, the active-site-saturation mutagenesis results support the design model, with the Tyr14–His98 and Asp18–Arg65 dyads having key roles in catalysis and the substrate-binding pocket largely conserved.

The most active catalysts, LuxSit-i (Extended Data Fig. 3b,e,h) and LuxSit-f (Extended Data Fig. 3c,f,i), were both expressed solubly in *E. coli* at high levels and are monomeric (some dimerization was observed at the high protein concentration; Extended Data Fig. 3l) and thermostable (Extended Data Fig. 3j,k). Similar to native luciferases that use CTZ, the apparent Michaelis constants (K_m) of both LuxSit-i and LuxSit-f are in the low-micromolar range (Fig. 3a) and the luminescent signal decays over time owing to fast catalytic turnover (Extended Data Fig. 7a). LuxSit-i is a very efficient enzyme, with a catalytic efficiency (k_{cat}/K_m) of 10^6 M^{-1} s^{-1}. The luminescence signal is readily visible to the naked eye (Fig. 3b), and the photon flux (photons per second) is 38% greater than that of the native *Renilla reniformis* luciferase (RLuc) (Supplementary Table 2). The DTZ luminescent reaction catalysed by LuxSit-i is pH-dependent (Extended Data Fig. 7b), consistent with the proposed mechanism. We used a combination of density functional theory (DFT) calculations and molecular dynamics (MD) simulations to investigate the basis for LuxSit activity in more detail; the results support the anion-stabilization mechanism (Extended Data Fig. 8a and Supplementary Fig. 3a) and suggest that LuxSit-i provides better DTZ transition-state charge stabilization than LuxSit (Extended Data Fig. 8b).

Fig. 3: Characterization of de novo luciferase activity in vitro and in human cells.


a
b
c
LuxSit-i
LuxSit-f
LuxSit
S/N = 13,888
S/N = 6,519
S/N = 124
Photons per second per molecule (×10⁻²)
0
10
20
30
40
0
10
20
30
40
50
Concentration (μM)
Fluorescence
Luminescence

a, Substrate concentration dependence of LuxSit, LuxSit-f and LuxSit-i activity. Numbers indicate the signal-to-background (*S*/*N*) ratio at V_{max} (photon s^{-1} molecule^{-1}). Data are mean ± s.d. ($n = 3$). **b**, Luminescence images acquired by a BioRad Imager (top) or an Apple iPhone 8 camera (bottom). Tubes from left to right: DTZ only; DTZ plus 100 nM purified LuxSit; and DTZ plus 100 nM purified LuxSit-i, showing the high efficiency of photon production. **c**, Fluorescence and luminescence microscopic images of live HEK293T cells transiently expressing LuxSit-i-mTagBFP2; LuxSit-i activity can be detected at single-cell resolution. Left, fluorescence channel representing the mTagBFP2 signal. Right, total luminescence photons were collected during a course of a 10-s exposure without excitation light, immediately after adding 25 µM DTZ. Insets, negative control, untransfected cells with DTZ. Scale bars, 20 µm; 40× magnification.

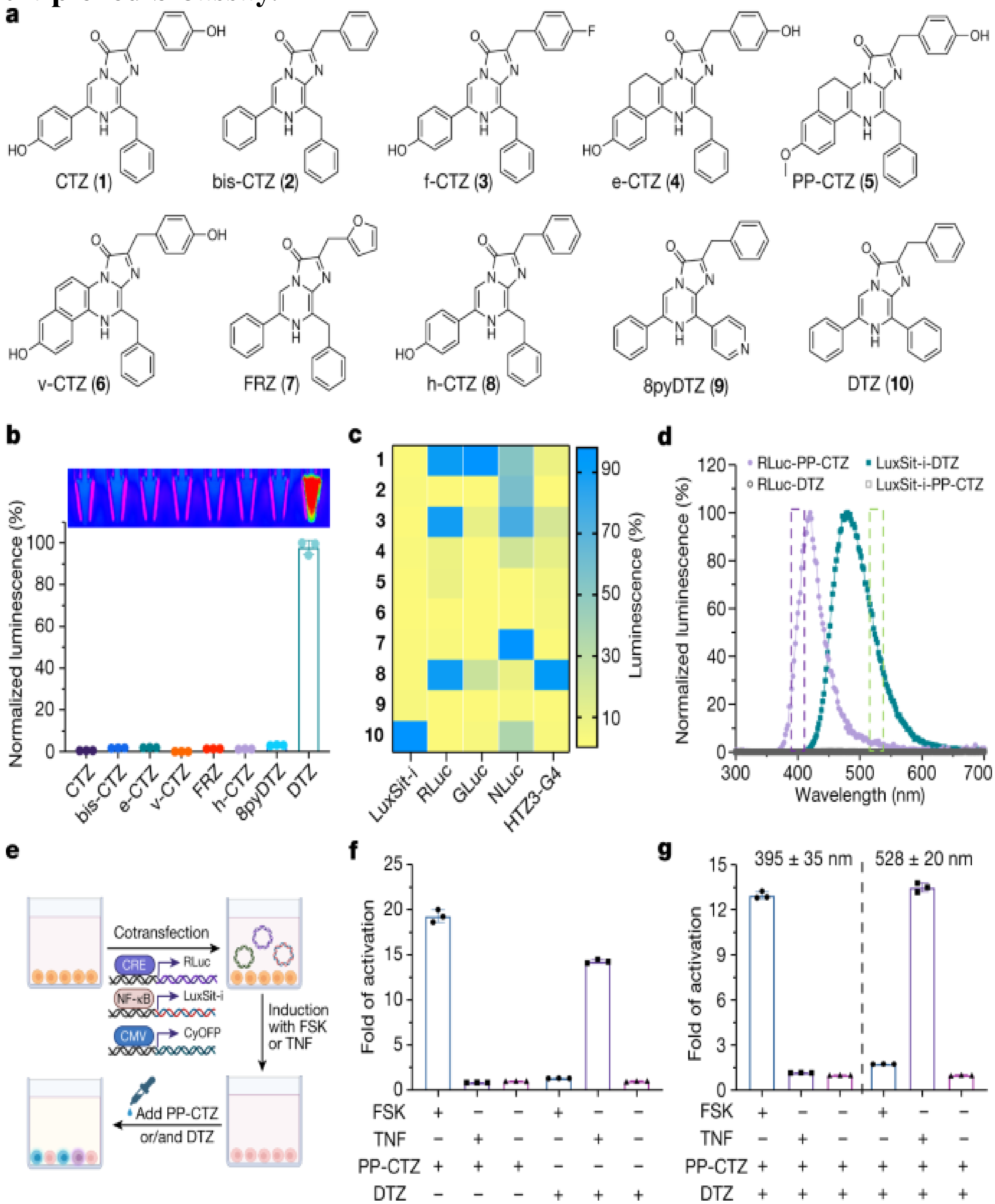
Source data

Cell imaging and multiplexed bioassay

As luciferases are commonly used genetic tags and reporters for cell biological studies, we evaluated the expression and function of LuxSit-i in live mammalian cells. HEK293T cells expressing LuxSit-i-mTagBFP2 showed DTZ-specific luminescence (Fig. 3c), which was maintained after targeting of LuxSit-i-mTagBFP2 to the nucleus, membrane and mitochondria (Extended Data Fig. 9). Native and previously engineered luciferases are quite promiscuous, with activity on many luciferin substrates (Fig. 4ac and Supplementary Fig. 4); this is possibly a result of their large and open pockets (a luciferase with high specificity to one luciferin substrate has been difficult to control even with extensive directed evolution[33,34]). By contrast, LuxSit-i exhibited exquisite specificity for its target luciferin, with 50-fold selectivity for DTZ over bis-CTZ (which differs only in one benzylic carbon; MD simulations suggest that this arises from greater transition-state shape complementarity (Extended Data Fig. 8b,c and Supplementary Fig. 3b,c)), 28-fold selectivity over 8pyDTZ (differing only in one nitrogen atom) and more than 100-fold selectivity over other luciferin substrates (Fig. 4b). One of our active design for h-CTZ (HTZ3-G4) was also highly specific for its target substrate (Fig. 4c and Extended Data Fig. 4d). Overall, the specificity of our designed luciferases is much greater than that of native

luciferases[35,36] or previously engineered luciferases[37] (Supplementary Table 5).

Fig. 4: High substrate specificity of de novo luciferases allows multiplexed bioassay.



a, Chemical structures of coelenterazine substrate analogues. **b**, Normalized activity of LuxSit-i on selected luciferin substrates. Luminescence image (top) and signal quantification (bottom) of the indicated substrate in the presence of 100 nM LuxSit-i. LuxSit-i has high specificity for the design target substrate, DTZ. **c**, Heat map visualization of the substrate specificity of LuxSit-i; *Renilla* luciferase (RLuc); *Gaussia* luciferase (GLuc); engineered NLuc from *Oplophorus* luciferase; and the de novo luciferase (HTZ3-G4) designed for h-CTZ. The heat map shows the luminescence for each enzyme on each substrate; values are normalized on a per-enzyme basis to the highest signal for that enzyme over all substrates. **d**, The luminescence emission spectrum of LuxSit-i-DTZ (green) and RLuc-PP-CTZ (purple) can be spectrally resolved by 528/20 and 390/35 filters (shown in dashed bars) and only recognize the cognate substrate. **e**, Schematic of the multiplex luciferase assay. HEK293T cells transiently transfected with CRE-RLuc, NF-κB-LuxSit-i and CMV-CyOFP plasmids were treated with either forskolin (FSK) or human tumour necrosis factor (TNF) to induce the expression of labelled luciferases. **f**,**g**, Luminescence signals from cells can be measured under either substrate-resolved or spectrally resolved methods by a plate reader. **f**, For the substrate-resolved method, luminescence intensity was recorded without a filter after adding either PP-CTZ or DTZ. **g**, For the spectrally resolved method, both PP-CTZ and DTZ were added, and the signals were acquired using 528/20 and 390/35 filters simultaneously. In **f** and **g**, the bottom panel indicates the addition of FSK or TNF. Luminescence signals were acquired from the lysate of 15,000 cells in CelLytic M reagent, and the CyOFP fluorescence signal was used to normalize cell numbers and transfection efficiencies. All data were normalized to the corresponding non-stimulated control. Data are mean ± s.d. ($n = 3$).

Source data

We reasoned that the high substrate specificity of LuxSit-i could allow the multiplexing of luminescent reporters through substrate-specific or spectrally resolved luminescent signals (Fig. 4d and Extended Data Fig. 10a,b). To investigate this possibility, we placed LuxSit-i downstream of the NF-κB response element and RLuc downstream of the cAMP response element (Fig. 4e). The addition of activators (TNF) of the NF-κB signaling pathway resulted in luminescence when cells were incubated with DTZ,

while the luminescence of PP-CTZ (the substrate of RLuc) was observed only when the cAMP–PKA pathway was activated (Fig. 4f). Because DTZ and PP-CTZ emit luminescence at different wavelengths, they can in principle be combined and the two signals can be deconvoluted through spectral analysis. Indeed, we observed that activating the NF-κB signaling resulted in luminescence at the DTZ wavelength, while the addition of cAMP–PKA pathway activators (FSK) generated luminescence at the PP-CTZ wavelength, allowing us to simultaneously assess the activation of the two signaling pathways in the same sample with either cell lysates (Fig. 4g) or intact HEK293T cells (Extended Data Fig. 10c–e) by providing both substrates together. Thus, the high substrate specificity of LuxSit-i enables multiplexed reporting of diverse cellular responses.

Conclusion

Computational enzyme design has been constrained by the number of available scaffolds, which limits the extent to which catalytic configurations and enzyme–substrate shape complementarity can be achieved[14,15,16]. The use of deep learning to produce large numbers of de-novo-designed scaffolds here eliminates this restriction, and the more accurate RoseTTAfold (ref. [38]) and AlphaFold2 (ref. [31]) should enable protein scaffolds to be generated even more effectively through family-wide hallucination and other approaches[18,39]. The diversity of shapes and sizes of scaffold pockets enabled us to consider a range of catalytic geometries and to maximize reaction intermediate–enzyme shape complementarity; to our knowledge, no native luciferases have folds similar to LuxSit, and the enzyme has high specificity for a fully synthetic luciferin substrate that does not exist in nature. With the incorporation of three substitutions that provide a more complementary pocket to stabilize the transition state, LuxSit-i has higher activity than any previous de-novo-designed enzyme, with a k_{cat}/K_m ($10^6\ M^{-1}\ s^{-1}$) in the range of native luciferases. This is a notable advance for computational enzyme design, as tens of rounds of directed evolution were required to obtain catalytic efficiencies in this range for a designed retroaldolase, and the structure was remodelled considerably[40]; by contrast, the predicted differences in ligand–side-chain interactions between LuxSit

and LuxSit-i are very subtle (Supplementary Fig. 2b; achieving such high activities directly from the computer remains a challenge in computational enzyme design). The small size, stability and robust folding of LuxSit-i makes it well-suited in luciferase fusions to proteins of interest and as a genetic tag for capacity-limited viral vectors. On the basic science side, the small size, simplicity and high activity of LuxSit-i make it an excellent model system for computational and experimental studies of luciferase catalytic mechanism. Extending the approach used here to create similarly specific luciferases for synthetic luciferin substrates beyond DTZ and h-CTZ would considerably extend the multiplexing opportunities illustrated in Fig. 4 (particularly with the recent advances in microscopy[41]), and enable a new generation of multiplexed luminescent toolkits. More generally, our family-wide hallucination method opens up an almost unlimited number of scaffold possibilities for substrate binding and catalytic residue placement, which is particularly important when the reaction mechanism and how to promote it are not completely understood: many alternative structural and catalytic hypotheses can be readily enumerated with shape and chemically complementary binding pockets but different catalytic residue placements. Although luciferases are unique in catalysing the emission of light, the chemical transformation of substrates into products is common to all enzymes, and the approach developed here should be readily applicable to a wide variety of chemical reactions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data for Figs. 2–4 are available online. The gene sequence for LuxSit-i has been deposited to GenBank under the accession number OP820699. See Supplementary Data for design models of LuxSit and LuxSit-i. Codon-optimized plasmids encoding LuxSit-i for bacterial and mammalian expression are available through Addgene. Source data are provided with this paper.

Code availability

The Rosetta macromolecular modelling suite (https://www.rosettacommons.org) is freely available to academic and non-commercial users. Commercial licences for the suite are available through the University of Washington Technology Transfer Office. The source code for RIF docking implementation is freely available at https://github.com/rifdock/rifdock. All relevant scripts and an accompanying Jupiter notebook for family-wide hallucination scaffold generation are available here: https://files.ipd.uw.edu/pub/luxSit/scaffold_generation.tar.gz. All generated scaffolds are available here: https://files.ipd.uw.edu/pub/luxSit/scaffolds.tar.gz. Computational design scripts for the luciferase libraries are available here: https://files.ipd.uw.edu/pub/luxSit/luciferase_designs_methods.zip.

References

1. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).

2. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).

3. Yeh, H. W. et al. Red-shifted luciferase–luciferin pairs for enhanced bioluminescence imaging. *Nat. Methods* **14**, 971–974 (2017).

4. Love, A. C. & Prescher, J. A. Seeing (and using) the light: recent developments in bioluminescence technology. *Cell Chem. Biol.* **27**, 904–920 (2020).

5. Syed, A. J. & Anderson, J. C. Applications of bioluminescence in biotechnology and beyond. *Chem. Soc. Rev.* **50**, 5668–5705 (2021).

6. Yeh, H.-W. & Ai, H.-W. Development and applications of bioluminescent and chemiluminescent reporters and biosensors. *Annu. Rev. Anal. Chem.* **12**, 129–150 (2019).

7. Zambito, G., Chawda, C. & Mezzanotte, L. Emerging tools for bioluminescence imaging. *Curr. Opin. Chem. Biol.* **63**, 86–94 (2021).

8. Markova, S. V., Larionova, M. D. & Vysotski, E. S. Shining light on the secreted luciferases of marine copepods: current knowledge and applications. *Photochem. Photobiol.* **95**, 705–721 (2019).

9. Jiang, T. Y., Du, L. P. & Li, M. Y. Lighting up bioluminescence with coelenterazine: strategies and applications. *Photochem. Photobiol. Sci.* **15**, 466–480 (2016).

10. Michelini, E. et al. Spectral-resolved gene technology for multiplexed bioluminescence and high-content screening. *Anal. Chem.* **80**, 260–267 (2008).

11. Rathbun, C. M. et al. Parallel screening for rapid identification of orthogonal bioluminescent tools. *ACS Cent. Sci.* **3**, 1254–1261 (2017).

12. Yeh, H.-W., Wu, T., Chen, M. & Ai, H.-W. Identification of factors complicating bioluminescence imaging. *Biochemistry* **58**, 1689–1697 (2019).

13. Su, Y. C. et al. Novel NanoLuc substrates enable bright two-population bioluminescence imaging in animals. *Nat. Methods* **17**, 852–860 (2020).

14. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De novo design of four-helix bundle metalloproteins: one scaffold, diverse reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).

15. Chino, M. et al. Artificial diiron enzymes with a de novo designed four-helix bundle structure. *Eur. J. Inorg. Chem.* **2015**, 3352–3352 (2015).

16. Basler, S. et al. Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold. *Nat. Chem.* **13**, 231–235 (2021).

17. Anishchenko, I. et al. De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).

18. Wang, J. et al. Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).

19. Norn, C. et al. Protein sequence design by conformational landscape optimization. *Proc. Natl Acad. Sci. USA.* **118**, e2017228118 (2021).

20. Yang, J. Y. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).

21. Basanta, B. et al. An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl Acad. Sci. USA* **117**, 22135–22145 (2020).

22. Loening, A. M., Fenn, T. D. & Gambhir, S. S. Crystal structures of the luciferase and green fluorescent protein from *Renilla reniformis*. *J. Mol. Biol.* **374**, 1017–1028 (2007).

23. Tomabechi, Y. et al. Crystal structure of nanoKAZ: the mutated 19 kDa component of *Oplophorus* luciferase catalyzing the bioluminescent reaction with coelenterazine. *Biochem. Biophys. Res. Commun.* **470**, 88–93 (2016).

24. Wu, N. et al. Solution structure of *Gaussia* luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear NMR. *Sci. Rep.* **10**, 20069 (2020).

25. Ding, B. W. & Liu, Y. J. Bioluminescence of firefly squid via mechanism of single electron-transfer oxygenation and charge-transfer-induced luminescence. *J. Am. Chem. Soc.* **139**, 1106–1119 (2017).

26. Isobe, H., Yamanaka, S., Kuramitsu, S. & Yamaguchi, K. Regulation mechanism of spin-orbit coupling in charge-transfer-induced luminescence of imidazopyrazinone derivatives. *J. Am. Chem. Soc.* **130**, 132–149 (2008).

27. Kondo, H. et al. Substituent effects on the kinetics for the chemiluminescence reaction of 6-arylimidazo[1,2-*a*]pyrazin-3(7*H*)-ones (*Cypridina* luciferin analogues): support for the single electron

transfer (SET)-oxygenation mechanism with triplet molecular oxygen. *Tetrahedron Lett.* **46**, 7701–7704 (2005).

28. Branchini, B. R. et al. Experimental support for a single electron-transfer oxidation mechanism in firefly bioluminescence. *J. Am. Chem. Soc.* **137**, 7592–7595 (2015).

29. Dou, J. Y. et al. De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).

30. Cao, L. et al. Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).

31. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

32. Dauparas, J. et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).

33. Yeh, H.-W. et al. ATP-independent bioluminescent reporter variants to improve in vivo imaging. *ACS Chem. Biol.* **14**, 959–965 (2019).

34. Xiong, Y. et al. Engineered amber-emitting nano luciferase and its use for immunobioluminescence imaging in vivo. *J. Am. Chem. Soc.* **144**, 14101–14111 (2022).

35. Bhaumik, S. & Gambhir, S. S. Optical imaging of *Renilla* luciferase reporter gene expression in living mice. *Proc. Natl Acad. Sci. USA* **99**, 377–382 (2002).

36. Szent-Gyorgyi, C., Ballou, B. T., Dagnal, E. & Bryan, B. Cloning and characterization of new bioluminescent proteins. In *Proc. SPIE 3600, Biomedical Imaging: Reporters, Dyes, and Instrumentation* (eds. Bornhop, D. J., Contag, C. H. & Sevick-Muraca, E. M.) https://doi.org/10.1117/12.351015 (SPIE, 1999).

37. Hall, M. P. et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).

38. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-+ (2021).

39. Wicky, B. I. M. et al. Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).

40. Giger, L. et al. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).

41. Yao, Z. et al. Multiplexed bioluminescence microscopy via phasor analysis. *Nat. Methods* **19**, 893–898 (2022).

Acknowledgements

We thank B. Wicky and R. Kibler for assistance on CD measurements; X. Li for help with mass spectrometry analysis of proteins; D. Feldman for the initial investigation of cell imaging; L. Milles for optimizing the golden gate assembly protocol; and the National Natural Science Foundation of China (22103060) for providing computational resources that were used in quantum mechanics and MD simulations. Some figure panels were partially created with BioRender.com (Fig. 4e and Extended Data Figs. 2 and 10c). Research reported in this publication was partially supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number K99EB031913. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We acknowledge funding from the Howard Hughes Medical Institute (Y.K., B.C. and D.B.); the Henrietta and Aubrey Davis Endowed Professorship in the University of Washington Department of Biochemistry (D.B.); the NIH Pathway to Independence Award (A.H.-W.Y., K99EB031913); the United World Antiviral Research Network (UWARN) (one of the Centers for Research in Emerging Infectious Diseases; CREIDs) NIAID 1 U01 AI151698-01 (D.B. and A.H.-W.Y.); the Audacious Project at the Institute for Protein Design (D.B. and A.H.-W.Y.); the Open Philanthropy Project Improving Protein Design Fund (D.B.); the Novo Nordisk Foundation (C.N., NNF18OC0030446), a Washington Research Foundation Fellowship (S.P.);

E. and W. Schmidt, by recommendation of the Schmidt Futures program (D.T., G.R.L., J.Z.Z., L.C., S.H., M.D., L.C. and D.B.); and the National Science Foundation (K.N.H., D.E. and P.M., CHE-1764328, and OCI-1053575 to XSEDE).

Author information

Author notes

1. These authors contributed equally: Andy Hsien-Wei Yeh, Christoffer Norn

Authors and Affiliations

1. Department of Biochemistry, University of Washington, Seattle, WA, USA

   Andy Hsien-Wei Yeh, Christoffer Norn, Yakov Kipnis, Doug Tischer, Samuel J. Pellock, Gyu Rie Lee, Jason Z. Zhang, Ivan Anishchenko, Brian Coventry, Longxing Cao, Justas Dauparas & David Baker

2. Institute for Protein Design, University of Washington, Seattle, WA, USA

   Andy Hsien-Wei Yeh, Christoffer Norn, Yakov Kipnis, Doug Tischer, Samuel J. Pellock, Gyu Rie Lee, Jason Z. Zhang, Ivan Anishchenko, Brian Coventry, Longxing Cao, Justas Dauparas, Samer Halabiya, Michelle DeWitt, Lauren Carter & David Baker

3. Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA

   Andy Hsien-Wei Yeh

4. Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

Yakov Kipnis, Brian Coventry & David Baker

5. Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA

Declan Evans, Pengchen Ma & K. N. Houk

6. School of Chemistry, Xi'an Key Laboratory of Sustainable Energy Materials Chemistry, MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, Xi'an Jiaotong University, Xi'an, China

Pengchen Ma

Contributions

D.B. supervised this work. A.H.-W.Y. conceived the project and performed the experimental characterization. A.H.-W.Y., Y.K. and C.N. conceptualized and investigated the initial design strategies and A.H.-W.Y. performed the computational design of luciferases. C.N. conceptualized and implemented the family-wide hallucination pipeline. C.N., D.T., S.J.P. and I.A. performed the scaffold design. S.J.P. performed the ProteinMPNN sequence design of NTF2 scaffolds. K.N.H. mentored D.E. and P.M. for MD and quantum mechanics calculations and contributed to the writing of the mechanistic parts of the manuscript. G.R.L. developed the computational SSM simulation. J.Z.Z. prepared mammalian cells and performed microscopy imaging. B.C. and L.C. integrated the tuning file function into RifDock. J.D. developed ProteinMPNN. S.H. helped assemble the designed library. M.D. and L.C. assisted with protein expression and purification. A.H.-W.Y., C.N. and D.B. wrote the initial manuscript. All authors discussed the results and contributed to the final manuscript.

Corresponding authors

Correspondence to Andy Hsien-Wei Yeh or David Baker.

Ethics declarations

Competing interests

A.H.-W.Y., C.N., Y.K., D.T., S.J.P., I.A. and D.B. are co-inventors in several provisional patent applications (application numbers 63/300171, 63/300178, 63/381922 and 63/381924 submitted by the University of Washington) covering the de novo luciferases and protein scaffolds described in this Article. A.H.-W.Y., C.N., J.Z. and D.B. are stockholders of Monod Bio, a company that aims to develop the inventions described in this manuscript. The remaining authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Proposed catalytic mechanism of coelenterazine-utilizing luciferases.

Density functional theory (DFT) calculation suggested that the formation of an anionic state is the essential electron source for the activation of triplet oxygen ($^3O_2$). Supported by both theoretical[25,26] and experimental evidence[27,28], the next oxygenation process is likely through a single-electron transfer (SET) mechanism in which the surrounding reaction field could highly influence the change of Gibbs free energy (ΔG_{SET}). Finally, the thermolysis of a dioxetane light emitter intermediate can produce photons via the mechanism of gradually reversible charge-transfer-induced

luminescence (GRCTIL), which is generally exergonic. As all the historical pieces of evidence are based on calculations in the virtual solvents or chemiluminescence in ideal organic solvents, the detailed mechanism of a luciferase-catalysed luminescence reaction has remained unclear. We proposed that the key step of the enzyme is to promote the formation of an anionic state and create a suitable environment to facilitate efficient SET. Hence, the goal of this study is to design an enzyme reaction field surrounding the substrate to stabilize the anionic substrate state and alter the local proton activity, solvent polarity, and hydrophobicity for the efficient activation of $^3O_2$.

Extended Data Fig. 2 Schematic representation of colony-based luciferase screening.

Computationally designed DNA sequences were purchased in an oligo array, where the fragments were amplified by PCR, assembled, and ligated into a pBAD bacterial expression vector. The plasmid library was used to transform DH10B cells. Each colony grown on the LB agar plate represented one luciferase design. The plates were sprayed with DTZ solution and imaged to identify active colonies using a ChemiDoc imager. Selected colonies were inoculated in 96-well plates, expressed, and purified to confirm individual luciferase activity. Plasmids can then be individually sequenced to point out active design models that provide insights into the design principle and enzyme functions or can be subjected to random mutagenesis for further evolution. Insert: three luciferases were identified from this screening. We refer to the most active and DTZ-specific luciferase as “LuxSit”.

Extended Data Fig. 3 Expression, purification and structural characterization of LuxSit variants.

a–**c**, The recombinant expression of **a**, LuxSit, **b**, LuxSit-i, and **c**, LuxSit-f in *E. coli*. Annotations for each lane are the following – 1: Pre-IPTG; 2: Post-IPTG; 3: Soluble lysate; 4: Flow-through; 5: Wash; 6: Elusion; 7: Post-TEV cleavage; 8: Post-SEC. **d**–**f**, Size-exclusion chromatography of the purified **d**, LuxSit; **e**, LuxSit-i; and **f**, LuxSit-f monomer. **g**–**i**, Deconvoluted mass spectrum of **g**, LuxSit, **h**, LuxSit-i, and **i**, LuxSit-f. **j**,**k**, Far-ultraviolet

circular dichroism (CD) spectra (Left panel) of **j**, LuxSit-i; and **k**, LuxSit-f at 25 °C (black line), 95 °C (red line) and cooled back to 25 °C (green line). CD melting curve at 220 nm (Right panel). **l**, Dimeric SEC peak was observed when LuxSit-i was concentrated to a high concentration (~50 μM) in Tris pH 8.0 buffer. Both dimeric and monomeric SEC fractions showed the expected size on SDS–PAGE and both peaks were catalytically active to emit luminescence in the presence of 25 μM DTZ.

Extended Data Fig. 4 Expression, purification and activity measurement of selected de-novo-designed luciferases for h-CTZ.

a, Coomassie-stained SDS–PAGE of HTZ3-D2 and HTZ3-G4 purified from recombinant expression in *E. coli*. **b**, Magnified views of HTZ3-D2 (left panel) and HTZ3-G4 (right panel) illustrated the side-chain preorganization of luciferase-h-CTZ interactions. **c**,**d**, Size-exclusion chromatography (left), deconvoluted mass spectrum (middle), and the normalized luciferase activities on selected compounds (right) of **c**, HTZ3-D2 and **d**, HTZ3-G4, which suggested high specificity for the design target substrate, h-CTZ. **e**, Substrate concentration dependence of LuxSit (w/ DTZ), HTZ3-D2 (w/ h-CTZ), and HTZ3-G4 (w/ h-CTZ) activity in PBS. All data points were fitted to the Michaelis-Menten equation. HTZ3-D2 and HTZ3-G4 showed K_m values of 7.9 and 19.5 μM with ~25% and ~58% I_{max} of LuxSit, respectively. Data are presented as mean ± s.d. ($n = 3$).

Extended Data Fig. 5 Predicted changes in substrate-binding free energy from binding-site mutations.

The calculated ddG_{bind} of each mutation was plotted as a function of the relative average experimental luciferase activity. The ddG_{bind} of hypothetical catalytic residues: **a**, Tyr14–His98 and **b**, Asp18–Arg65 dyads were generally not the lowest, which suggested that these designed catalytic residues are not favourable for substrate binding. Red dots represent the wild-type (LuxSit) amino acids. The rank of wild-type ddG_{bind} for each position screened for activity is shown with a heat map in **c**. **d**–**f**, The wild-type ddG_{bind} of the residues designed for **d**,**e**, π–π stacking or **f**, hydrophobic

interactions were the lowest compared to the mutation ddG_{bind} values. This shows that the sequence is near-optimal for substrate binding and the design model is reliable.

Extended Data Fig. 6 Screening of a randomized NNK library at 60, 96 and 110 positions and sequence alignment between LuxSit and its variants.

We generated a fully randomized library at 60, 96, and 110 positions to screen all possible combinations exhaustively. After the colony-based screening, we identified many colonies with strong luciferase activities with DTZ. Each colony was expressed individually in each well of 96-well plates (1 mL culture) and purified accordingly (see Supplementary Methods). **a**, Individual luminescence activity of each selected mutant was plotted and compared to the parent, LuxSit. Luminescence activities were measured in the presence of 25 μM DTZ. Luminescence activity (RLU) was shown as the integrated signal over the first 15 min. Statistical analysis of the amino acid frequency versus the luciferase activity at residue **b**, 60, **c**, 96, and **d**, 110. Data are presented as mean ± s.d. (n varies across each bar as the mutants were selected from a randomized library). Arg60 is confirmed to be mutable among all selected mutants as Arg60 may be structurally less well-defined because it emanates from a loop and has no hydrogen-bonding partner. Ala96 prefers larger side-chain substitutions (Leu, Ile, Met, and Cys), and Met110 favours hydrophobic residues (Val, Ile, and Ala). A newly discovered variant (R60S/A96L/M110V) with more than 100-fold higher photon flux over LuxSit was assigned LuxSit-i for its high brightness. In the sequence alignment, mutations are highlighted in yellow fonts and grey backgrounds. The conserved catalytic dyads of Asp18–Arg65 and Tyr14–His98 are in green and blue fonts.

Extended Data Fig. 7 Additional characterization of LuxSit variants.

a, Normalized emission kinetics of 15,000 intact HeLa cells expressing LuxSit-i (red), 100 nM purified LuxSit-i (green), or 100 nM purified LuxSit-f (blue) in the presence of 50 μM DTZ. The more extended emission kinetics

in HeLa cells is likely due to the diffusion rate of DTZ across cell membranes. **b**, Normalized luminescence decay curves of LuxSit-i in various pH buffers revealed a pH-dependent catalytic mechanism. **c**, Luminescent quantum yield was estimated from the integrated luminescence signal until completely converting 125 pmol substrates to photons in the presence of 50 nM corresponding luciferase (see Supplementary Methods). Data are presented as mean (*n* = 3).

Extended Data Fig. 8 Free-energy profile of DTZ chemiluminescence and MD simulations of proposed protein–intermediate complexes.

a, The free-energy profile calculated by density functional theory (DFT) shows triplet oxygen can react directly with the anionic species of DTZ (**Int1**) through the reactant complex **Int2** and **TS1**. The dioxetane intermediate **Int3** then cleaves in an open shell singlet transition state $^{\mathbf{OSS}}$**TS2** to form excited intermediate **Int4***, which rapidly extrudes CO_2 and forms the emissive product **Int5**. Note: either **Int4*** or **Int5*** emit in the observed region, but the lifetime of **Int4*** is very short and likely completely converts to **Int5*** before emission. **b**, **Int2** and **Int3** were docked into both LuxSit and LuxSit-i and the bindings were evaluated by molecular dynamics (MD). The distances between His98 to O1 (top row) and Arg65 to N1 (bottom row) of the substrate were plotted throughout 500 ns MD simulations. LuxSit-i (blue trace) binds **Int2′** (middle) considerably better than LuxSit does (red trace), suggesting that the mutations of LuxSit-i provide a binding pocket more complimentary to **TS1**. This binding orientation brings N1 of the substrate much closer to Arg65, providing better charge stabilization for the high energy transition state. **c**, Docking of the peroxide anion form of bis-CTZ into the pocket of LuxSit-i; blue overlay represents DTZ in the original design model. During MD simulation, the added benzylic carbon of bis-CTZ (green trace) disrupts the shape complementarity between LuxSit-i and the transition states (**TS1** and **TS2**), reducing the charge stabilization by Arg65. This charge stabilization is necessary for the reaction to proceed, explaining the high substrate specificity of LuxSit-i for DTZ over bis-CTZ.

Extended Data Fig. 9 Expression, localization and luminescence activity of LuxSit-i in live HEK293T and HeLa cells.

a,**b**, Fluorescence imaging of live **a**, HEK293T and **b**, HeLa cells expressing LuxSit-i-mTagBFP2, which is untargeted or localized to the nucleus (Histone2B), plasma membrane (KRasCAAX), or mitochondria (DAKAP) cellular compartments. Scale bar: 10 µm. **c**,**d**, Luminescence signals were measured with 15,000 intact **c**, HEK293T or **d**, HeLa cells in the presence of 25 µM DTZ in DPBS. Transfection efficiencies range from 60-70% for HEK293T cells and 5-10% for HeLa cells. **e**, Luminescence emission spectra acquired from LuxSit-i expressing HEK293T cells is consistent with the emission spectra of recombinant LuxSit-i purified from *E. coli*. **f**,**g**, Luminescence signals were measured with 15,000 **f**, intact LuxSit-i expressing HEK293T cells or **g**, cell lysate in the presence of 25 µM indicated substrate in DPBS. Luminescence intensities were normalized to DTZ signal, showing high DTZ specificity over other substrates in cell-based assays. Data were shown as total luminescence signal over the first 20 min ± s.d. ($n = 3$). **h**, Normalized luminescence intensity profile of lines traversing across different cells ($n = 10$) of main Fig. 3c luminescence image; grey lines represent untransfected cells. Error bars represent ± SEM.

Extended Data Fig. 10 Substrate specificity of LuxSit-i and spectrally resolved luciferase–luciferin pairs allow multiplexed bioassay.

a, The orthogonality relationship between LuxSit-i-DTZ and RLuc-PP-CTZ (Prolume Purple, methoxy e-Coelenterazine) luminescent pairs. Indicated percentages of each luciferase were mixed at different ratios totalling 100%. After the addition of both 25 µM DTZ and PP-CTZ substrates, filtered light from 528/20 and 390/35 channels were measured simultaneously. **b**, Heat map shows the luminescence signal for individual luciferase (100 nM) or 1:1 mixture in the presence of the cognate or non-cognate (DTZ or PP-CTZ or both) substrates. Response signals were acquired by a Neo2 plate reader with 528/20 and 390/35 nm filters simultaneously. **c**, Multiplex luciferase assay in live HEK293T after co-transfection of CRE-RLuc, NFκB-LuxSit-i, and CMV-CyOFP plasmids and stimulation by Forskolin (FSK) or human TNF. **d**,**e**, 15,000 intact cells were assayed (see Supplementary Methods) by

either **d**, substrate-resolved or **e**, spectrally resolved modes after adding DTZ, PP-CTZ, or both DTZ and PP-CTZ in DPBS without cell lysis. Area scanning of the CyOFP fluorescence signal was used to estimate cell numbers and transfection efficiency. The reported unit was RLU/a.u.; relative light units/fluorescence intensity measurements at Ex./Em. = 480/580 nm. All data were normalized to the corresponding non-stimulated control. Data are presented as mean ± s.d. ($n = 3$).

Supplementary information

Supplementary Information

This file contains the Methods section

Reporting Summary

Peer Review File

Design models of LuxSit and LuxSit-I

Source data

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The

images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Yeh, A.HW., Norn, C., Kipnis, Y. *et al.* De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023). https://doi.org/10.1038/s41586-023-05696-3

- Received: 19 January 2022
- Accepted: 03 January 2023
- Published: 22 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05696-3

Share this article

Anyone you share the following link with will be able to read this content:

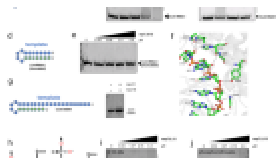
Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

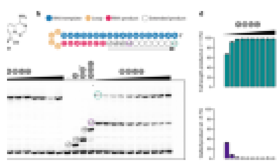
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05696-3



Characterization of SARS-CoV-2 replication complex elongation and proofreading activity

10 June 2022

Alisha N. Jones, André Mourão, … Grzegorz M. Popowicz



Mechanism of SARS-CoV-2 polymerase stalling by remdesivir

12 January 2021

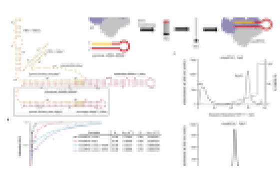
Goran Kokic, Hauke S. Hillen, … Patrick Cramer



Architecture of an HIV-1 reverse transcriptase initiation complex

25 April 2018

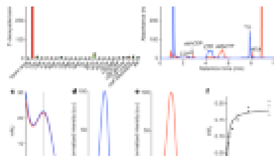
Kevin P. Larsen, Yamuna Kalyani Mathiharan, … Elisabetta Viani Puglisi



[High-resolution view of HIV-1 reverse transcriptase initiation complexes and inhibition by NNRTI drugs]

04 May 2021

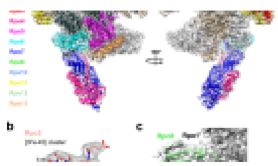
Betty Ha, Kevin P. Larsen, … Elisabetta Viani Puglisi



[A naturally occurring antiviral ribonucleotide encoded by the human genome]

20 June 2018

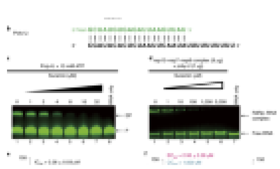
Anthony S. Gizzi, Tyler L. Grove, … Steven C. Almo



[Structural basis of RNA polymerase inhibition by viral and host factors]

17 September 2021

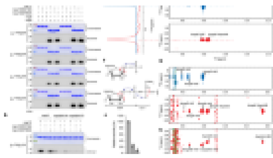
Simona Pilotto, Thomas Fouqueau, … Finn Werner



[Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin]

05 March 2021

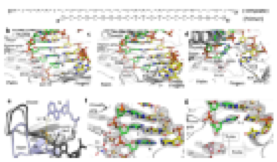
Wanchao Yin, Xiaodong Luan, … H. Eric Xu



The mechanism of RNA capping by SARS-CoV-2

09 August 2022

Gina J. Park, Adam Osinski, … Vincent S. Tagliabracci



Sliding of HIV-1 reverse transcriptase over DNA creates a transient P pocket – targeting P-pocket by fragment screening

08 December 2021

Abhimanyu K. Singh, Sergio E. Martinez, … Kalyan Das

- Article
- Published: 01 February 2023

Structural basis for substrate selection by the SARS-CoV-2 replicase

- Brandon F. Malone,
- Jason K. Perry,

- Paul Dominic B. Olinares,
- Hery W. Lee,
- James Chen,
- Todd C. Appleby,
- Joy Y. Feng,
- John P. Bilello,
- Honkit Ng,
- Johanna Sotiris,
- Mark Ebrahim,
- Eugene Y. D. Chua,
- Joshua H. Mendez,
- Ed T. Eng,
- Robert Landick,
- Matthias Götte,
- Brian T. Chait,
- Elizabeth A. Campbell &
- …
- Seth A. Darst

Show authors

Nature **volume 614**, pages 781–787 (2023)

- 4627 Accesses
- 21 Altmetric
- Metrics details

Abstract

The SARS-CoV-2 RNA-dependent RNA polymerase coordinates viral RNA synthesis as part of an assembly known as the replication–transcription complex (RTC)[1]. Accordingly, the RTC is a target for clinically approved antiviral nucleoside analogues, including remdesivir[2]. Faithful synthesis of viral RNAs by the RTC requires recognition of the correct nucleotide triphosphate (NTP) for incorporation into the nascent RNA. To be effective

inhibitors, antiviral nucleoside analogues must compete with the natural NTPs for incorporation. How the SARS-CoV-2 RTC discriminates between the natural NTPs, and how antiviral nucleoside analogues compete, has not been discerned in detail. Here, we use cryogenic-electron microscopy to visualize the RTC bound to each of the natural NTPs in states poised for incorporation. Furthermore, we investigate the RTC with the active metabolite of remdesivir, remdesivir triphosphate (RDV-TP), highlighting the structural basis for the selective incorporation of RDV-TP over its natural counterpart adenosine triphosphate[3,4]. Our results explain the suite of interactions required for NTP recognition, informing the rational design of antivirals. Our analysis also yields insights into nucleotide recognition by the nsp12 NiRAN (nidovirus RdRp-associated nucleotidyltransferase), an enigmatic catalytic domain essential for viral propagation[5]. The NiRAN selectively binds guanosine triphosphate, strengthening proposals for the role of this domain in the formation of the 5′ RNA cap[6].

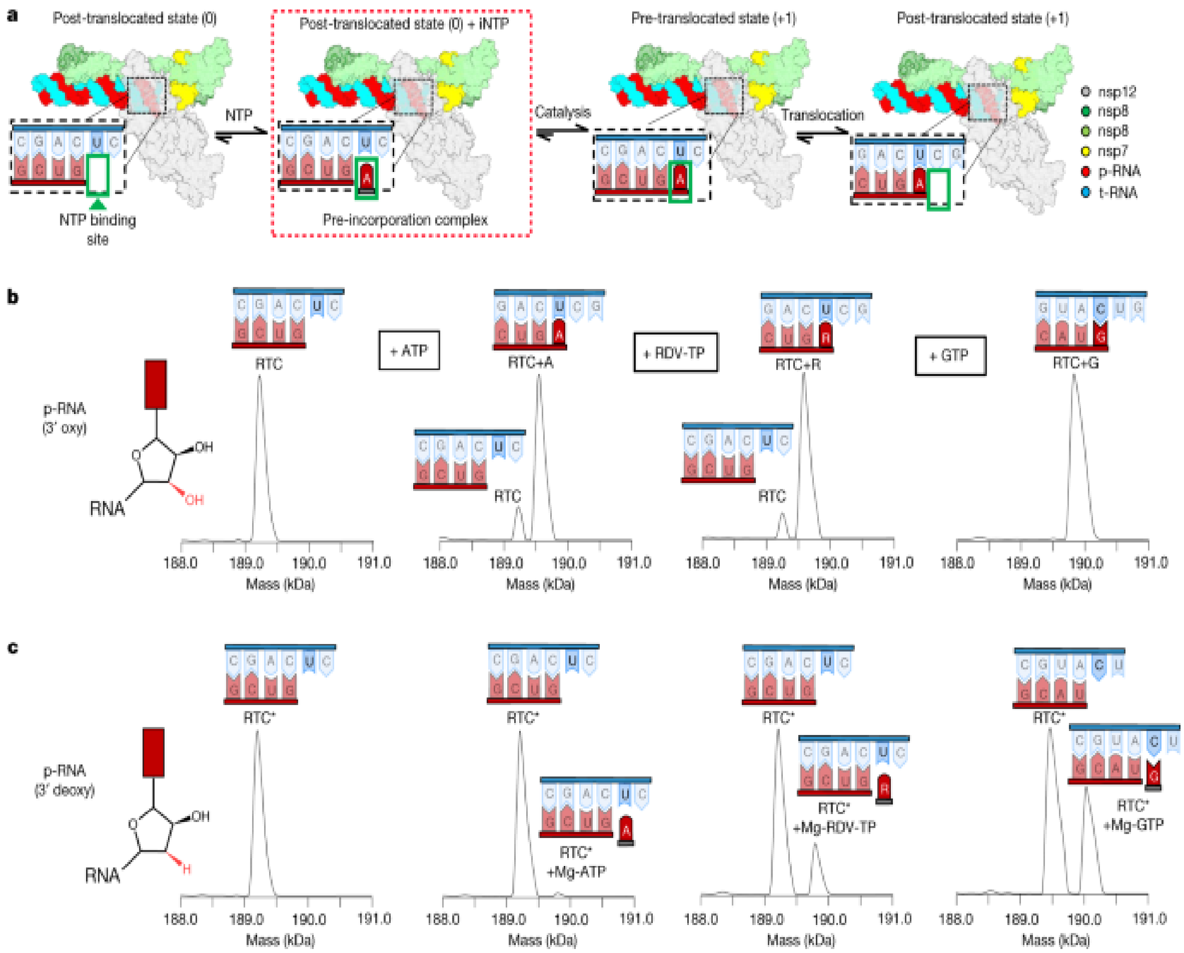
Main

COVID-19, caused by the coronavirus SARS-CoV-2, continues to devastate livelihoods and overwhelm healthcare systems worldwide. Given the urgency of stymieing infection and mitigating disease morbidity, a concerted research effort has begun to explain the molecular details of the viral life cycle and to design therapeutics to disrupt it[1]. The viral RNA-dependent RNA polymerase (RdRp, encoded by non-structural protein 12, or nsp12) functions as part of the holo-RdRp (comprising the RdRp and auxiliary proteins nsp7 and nsp8 as a heterotetramer nsp7/nsp8$_2$/nsp12) in a replication–transcription complex (holo-RdRp + RNA, or RTC) to direct all viral RNA synthesis in conjunction with a group of viral nucleic acid processing enzymes. In less than 2 years, targeting the RTC by the antivirals remdesivir (RDV) and molnupiravir has become an important aspect of clinical care[2,7]. This clinical success underscores the importance of the RTC as a pharmacological target and incentivizes the design of more efficacious nucleotide analogues for the treatment of COVID-19. Furthermore, because of the highly conserved nature of the coronavirus RdRp active site[8], nucleotide analogue inhibitors of the RTC are excellent candidates for pan-

viral inhibitors that would be effective against emerging variants and may be repurposed to tackle future pathogenic coronaviruses that could arise.

Biochemical and single-molecule experiments have characterized the RNA synthesis activity of the SARS-CoV-2 RTC, its selectivity for a wide range of nucleotide analogues and their effects on RNA synthesis[3,4,9,10]. These investigations yielded insights into the ability of the RTC to discriminate between natural NTPs and related antiviral analogues. Studies of remdesivir (RDV) revealed that the active metabolite of RDV, RDV triphosphate (RDV-TP), possesses enhanced selectivity for incorporation into nascent RNA over its natural counterpart, adenosine triphosphate (ATP)[3,4]. Rapid, competitive incorporation is probably a critical facet of the clinical effectiveness of RDV. However, the structural basis for NTP recognition and RDV-TP selectivity remains unknown, necessitating structural examinations of stalled RTC ternary complexes with incoming NTP substrates and catalytic metal ions poised for catalysis (Michaelis or pre-incorporation complexes) (Fig. 1a).

Fig. 1: Capturing the SARS-CoV-2 RTC ternary complex.



a, Schematic depicting the major steps of the nucleotide addition cycle of the RTC. The pre-incorporation complex studied here is highlighted (red dashed box). **b**, nMS analysis of the RTC bound to a 3′ oxy p-RNA in the absence and presence of 300 µM ATP, RDV-TP or GTP, respectively. **c**, Similar nMS analysis to **b** but the RTC was reconstituted using a 3′-deoxy p-RNA (RTC*). p-RNA and t-RNA schematics were created with BioRender.com.

Stepwise nucleotide incorporation necessitates transferring the 3′-end of the product-RNA (p-RNA) to the post-translocated site, allowing binding of the incoming NTP substrate along with two catalytic metal ions in the active-site cleft (Fig. 1a)[11,12,13,14]. In the subsequent catalytic step(s), the geometry of correct Watson–Crick base pairing of the incoming NTP with the template-RNA (t-RNA) base closes the active site. This conformational change promotes nucleophilic attack of the p-RNA 3′-OH on the NTP α-phosphate, releasing a pyrophosphate in a S_N2 condensation reaction[11].

Available SARS-CoV-2 RTC structures are dominated by apo structures with an empty NTP active site lacking catalytic metals and with an open active-site conformation, hindering structure-based drug design[15,16,17,18]. Only two studies document the RTC in the presence of substrates: the antiviral inhibitors favipiravir triphosphate and the activated form of AT-527, AT-9010, but neither investigation yielded views of a catalytically competent RTC because of unproductive binding modes and open conformations of the active sites[19,20]. Therefore, to gain mechanistic insights into RTC recognition of its substrates during RNA synthesis, we determined five cryogenic-electron microscopy (cryo-EM) structures, capturing the RTC with each of the four natural NTPs as well as with RDV-TP. The five cryo-EM maps range in nominal resolutions between roughly 2.6 and 3.3 Å, with the RdRp active sites resolved locally to roughly 2.2–2.9 Å (Extended Data Figs. 1–5 and Supplementary Table 1)[21], enabling near-atomic resolution insights into the mechanism of NTP recognition and RDV-TP selectivity.

Trapping RTC pre-incorporation complexes

Guided by previous mechanistic work on nucleic acid polymerases, we investigated a battery of chemical strategies to block incorporation after substrate binding. These approaches included using nucleotide diphosphate (NDP) substrates[12], alternative metal cofactors (such as Ca^{2+})[22], α-β non-hydrolysable nucleotide analogues[23] and 3′-deoxy p-RNA scaffolds. Although the use of NDP substrates trapped the hepatitis C virus (HCV) RdRp recognition complex[12], the SARS-CoV-2 RTC incorporated guanosine diphosphate (GDP), adenosine diphosphate (ADP) and to a minor extent RDV-diphosphate (RDV-DP) under assay conditions similar to the cryo-EM sample preparation (Extended Data Fig. 6, see also ref. [24]). The ability of the SARS-CoV-2 RTC to incorporate NDPs reflects the relative promiscuity of the RdRp active site, which also retained RNA synthesis activity in the presence of both Ca^{2+} and α-β-imino analogues (Extended Data Fig. 6e). However, RTC synthesis activity was fully ablated with the incorporation of a 3′-deoxy nucleotide into the p-RNA, leading us to use this approach to visualize the RTC pre-incorporation complexes (Extended Data Fig. 6g).

To validate the effectiveness of using 3′-deoxy RNA scaffolds, we used native mass spectrometry (nMS) to probe for the extension of the primer by one base or the formation of a stalled ternary complex (Fig. 1b,c and Supplementary Fig. 2) in the presence of various NTPs. In the event of nucleotide incorporation, we expected a mass-shift corresponding to nucleotide monophosphate addition, whereas non-covalently bound (unincorporated) substrates would either dissociate or yield a mass-shift corresponding to the NTP and bound metal ions. For the control scaffolds containing a 3′-OH, near complete p-RNA extension by a single nucleotide monophosphate occurred (Fig. 1b). In the presence of p-RNA strands lacking a 3′-OH, no incorporation of the next nucleotide (ATP, RDV-TP or guanosine triphosphate (GTP)) was observed, confirming the suitability of this strategy for trapping pre-incorporation complexes (Fig. 1c). In the pre-incorporation complexes, we observed peaks corresponding to the RTC with the incoming Mg-NTP (ATP, RDV-TP or GTP) with the relative peak intensity in the order of GTP > RDV-TP > ATP (Fig. 1c).

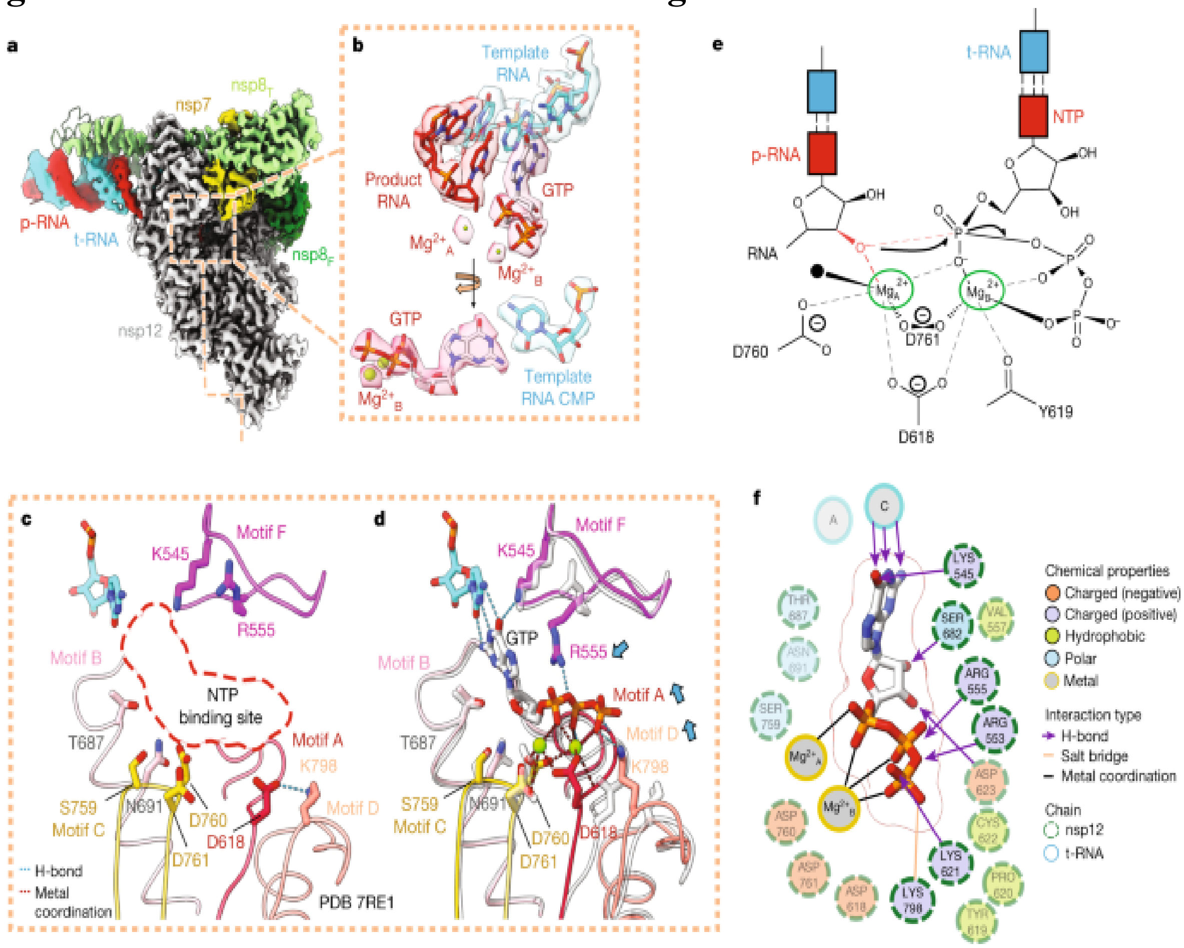
Structural basis of NTP recognition

Having identified a strategy to capture substrate-bound RTCs stalled in a pre-incorporation state, we prepared cryo-EM samples of RTCs with each of the natural nucleotides, ATP, cytidine triphosphate (CTP), GTP, uridine triphosphate (UTP) or with RDV-TP (Extended Data Figs. 1–5 and Supplementary Table 2). The resulting structures yielded the first views of the RTC poised for catalysis. Two of the structures, the RTC with CTP (+CTP) or GTP (+GTP), yielded nominal resolutions better than 3 Å (2.7 Å), with local resolution around the active site estimated to be roughly 2.2 Å (ref. [21]), enabling near-atomic resolution insights into substrate recognition.

Density features in the +CTP and +GTP cryo-EM maps are consistent with the positions of divalent metal ions observed in many previous nucleic acid polymerase structures[11,12,13,25]. On the basis of the presence of 5 mM $MgCl_2$ in the cryo-EM buffer and the observed octahedral coordination geometries, we assigned these densities as $Mg^{2+}{}_A$ and $Mg^{2+}{}_B$ (Fig. 2b and Supplementary Video 1), consistent with the hypothesized universal two-metal ion mechanism for nucleic acid polymerases[11]. The density peak for

$Mg^{2+}{}_A$ in the +CTP and +GTP structures is weaker than that for $Mg^{2+}{}_B$, and density for $Mg^{2+}{}_A$ in the +ATP, +UTP and +RDV-TP structures is completely absent. Despite the absence of $Mg^{2+}{}_A$ in the +ATP, +UTP and +RDV-TP structures, most details of the active-site configurations were indistinguishable from the +CTP and +GTP structures.

Fig. 2: Structural basis of nucleotide recognition.



a, Cryo-EM density of the 2.7 Å nominal resolution +GTP structure (S4_GTP), coloured according to the fitted model chains. **b**, Cryo-EM density of the bound incoming GTP, two associated metal ions and the nearby bases of the t-RNA and p-RNA strands. **c**, Close-up of the active site of the apo complex PDB 7RE1 (ref. [18]), illustrating the arrangement of the conserved RdRp active site motifs A–D and F (crimson, hot pink, gold, salmon and magenta, respectively) and of key residues when the NTP binding site (dashed red line) is empty. **d**, Comparison of the S4_GTP and

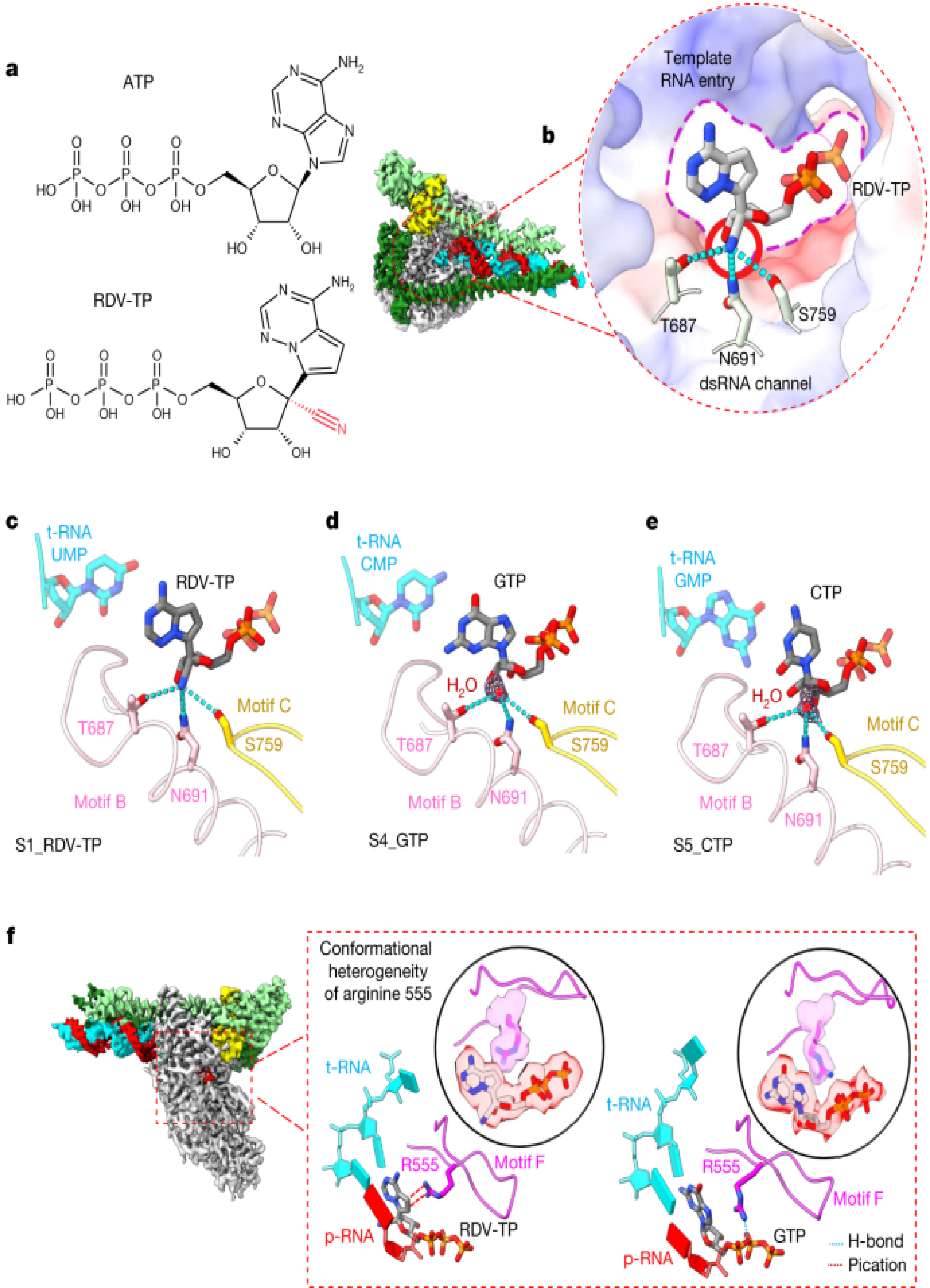
apo structure (faded grey) active sites, highlighting observed motif/residue rearrangements (blue arrows) on NTP binding. Movements in motifs A and D close the active site. **e**, Schematic depicting nucleotide addition, in the presence of an intact 3′-OH, based on disposition of magnesiums and NTP in the S4_GTP structure. **f**, 2D schematic of the suite of interactions involved in NTP binding based on distances in the S4_GTP structure. Residues interacting indirectly or weakly with the NTP are faded for clarity.

A consistent feature observed across all five structures is the closure of the active site around the incoming NTP, mediated by a rotation of the RdRp motif A towards the NTP substrate (Fig. 2c,d). Closure of motif A stabilizes substrate binding by: (1) enabling the backbone carbonyl of Y619 to coordinate $Mg^{2+}{}_B$ and the catalytic residue D618 to coordinate both $Mg^{2+}{}_A$ and $Mg^{2+}{}_B$ (Fig. 2e), (2) promoting the formation of a hydrogen-bonding (H-bonding) network through D623 that enables the recognition of the substrate ribose 2′-OH by S682, T687 and N691 (motif B), (3) enabling weak H-bonding interactions between the β- and γ-phosphates and motif A residues K621 or C622 (Fig. 2f, and Extended Data Fig. 7f) and (4) disrupting the polar D618–K798 interaction in the apo-RTC[15,16,18], repositioning D618 for metal coordination and K798 to interact with the NTP γ-phosphate (Fig. 2d).

The detailed interactions of nsp12 residues with the four natural NTP substrates were essentially identical except for conserved basic residues of motif F. K545 (motif F) shows selective interactions with the incoming NTPs, forming a H-bond with carbonyl oxygens of UTP (C4) and GTP (C6), interactions that are precluded by the amino group at these positions in the ATP, RDV-TP or CTP bases (Extended Data Fig. 7). The cryo-EM densities also suggest that the side chains of K551 and R555 are dynamic, accessing several conformations in each structure. K551 primarily interacts with the γ-phosphate of the CTP and RDV-TP substrates, but weak density indicates disorder in the other structures. In the +ATP and +RDV-TP structures, R555 predominantly forms a pi–pi stacking interaction with the NTP substrate base. This differs from the predominant conformation in the +CTP, +GTP and +UTP structures where the side chain of R555 forms a H-bond with the β-phosphate oxygens of the NTP substrate (Fig. 3 and Extended Data Fig. 7). In either state, R555 buttresses the incoming nucleotide, arranging the

active site for catalysis. We speculate that the observed R555 stacking interaction may first promote the formation of a canonical Watson–Crick base pair by stabilizing the nucleobase opposite the template base, then reorient to interact with the nucleotide β-phosphate to promote subsequent catalysis by reinforcing the correct geometry of the pyrophosphate leaving group, roles that may be conserved across RdRps such as HCV RdRp[26].

Fig. 3: Molecular basis of remdesivir's incorporation selectivity.


a
ATP
RDV-TP
b
Template RNA entry
RDV-TP
T687
N691
S759
dsRNA channel
c
t-RNA UMP
RDV-TP
Motif C
T687
S759
Motif B
N691
S1_RDV-TP
d
t-RNA CMP
GTP
H2O
Motif C
T687
S759
Motif B
N691
S4_GTP
e
t-RNA GMP
CTP
H2O
Motif C
T687
S759
Motif B
N691
S5_CTP
f
Conformational heterogeneity of arginine 555
t-RNA
R555
Motif F
p-RNA
RDV-TP
t-RNA
R555
Motif F
GTP
p-RNA
H-bond
Pication

a, Chemical structures of ATP and RDV-TP, highlighting the position of the RDV-TP 1′ cyano group. **b**, Cryo-EM densities of the S2_RDV-TP structure, coloured according to fitted model chains. Zoom-in on the bound RDV-TP illustrates the RDV-TP 1′ cyano group is accommodated in a hydrophilic pocket formed by motif B and C residues. Protein surface is coloured according to electrostatics. **c**–**e**, Comparison of the active sites of the S1_RDV-TP (**c**), S4_GTP (**d**) and S5_CTP (**e**) structures reveals that the RdRp cyano pocket can also bind a water molecule (map density around the water shown in mesh), which needs to be displaced for RDV-TP binding. **f**, A comparison of the S2_RDV-TP and S4_GTP structures reveals two predominant rotamers of R555 that mediate either a pi–pi stacking (red dotted lines) or a H-bond (blue dotted line) interaction with the incoming NTP.

The structural observation of SARS-CoV-2 RdRp motif A closure resembles a similar motion of motif A observed in substrate-bound RdRp complexes of polio virus, norovirus and HCV[12,13,14], illustrating the universal nature of the RdRp palm closure step during NTP recognition. Although the palm-mediated active site closure is well-documented, our results indicate that the structural plasticity of positively charged residues in motif F (K545, K551, R553 and R555) in the fingers domain helps orient the NTP substrate for catalysis through interactions with both the nucleobase and triphosphate moieties. We note that these residues are invariant across the major coronavirus clades and GISAID database of patient SARS-CoV-2 sequences (Extended Data Fig. 8), reflecting the near immutable nature of the coronavirus RdRp active site. Previously, we noted that the same residues line the NTP entry channel following nsp13-mediated p-RNA backtracking[27], illustrating the pleiotropic role that motif F plays during RNA synthesis.

Structural basis of RDV-TP selectivity

RDV-TP has been characterized to possess a twofold to threefold greater selectivity for incorporation into elongating RNA compared to ATP[3,4], a property that is thought to be critical for its clinical efficacy. In our structures, both nucleotides are observed base-paired to the cognate t-RNA U; the two nucleotides superimpose with a root-mean-square-deviation of

0.34 Å over 32 common atoms. The key difference between RDV-TP and ATP is the 1′-cyano moiety on the RDV-TP ribose (Fig. 3a), which juts into a hydrophilic pocket formed by residues T687, N691 (motif B) and S759 (motif C) (Fig. 3c), creating a network of polar interactions (Fig. 3c). In the highest resolution cryo-EM maps (+CTP and +GTP, Supplementary Table 1), we observed a stably bound water molecule occupying the hydrophilic pocket in the absence of the RDV-TP cyano group (compare Fig. 3c to 3d,e), suggesting that the enhanced affinity (lower K_m) for RDV-TP over ATP[3,4] may be attributed to both an entropic effect through release of the bound water and an enthalpic effect through formation of new polar interactions between RDV-TP and the surrounding polar residues.

A recent study reports that SARS-CoV-2 acquired phenotypic resistance to RDV through an nsp12-S759A substitution following serial passage in cell culture[28], indicating that binding of the RDV-TP 1′-cyano group in this hydrophilic pocket is an important facet of RDV susceptibility. To investigate whether the polar interactions observed in the 1′-cyano pocket are important for RDV-TP selectivity, we assayed to discover whether RDV-TP incorporation is perturbed in the presence of single alanine substitutions of S759, T687 or N691. Consistent with previous work[3,4], the wild-type nsp12 enzyme possesses an almost twofold greater selectivity for RDV-TP over ATP whereas the S759A mutation renders the enzyme fivefold more selective for ATP (Extended Data Fig. 9). Both the T687A and N691A mutations greatly impair RNA synthesis activity regardless of whether ATP or RDV-TP was the substrate, however, the mutant enzymes strongly prefer to incorporate ATP over RDV-TP (Supplementary Fig. 3). We conclude that accommodation of the RDV-TP 1′-cyano group in the hydrophilic pocket formed from nsp12 S759, T687 and N691 explains how RDV-TP outcompetes ATP as an RdRp substrate[3,4,28].

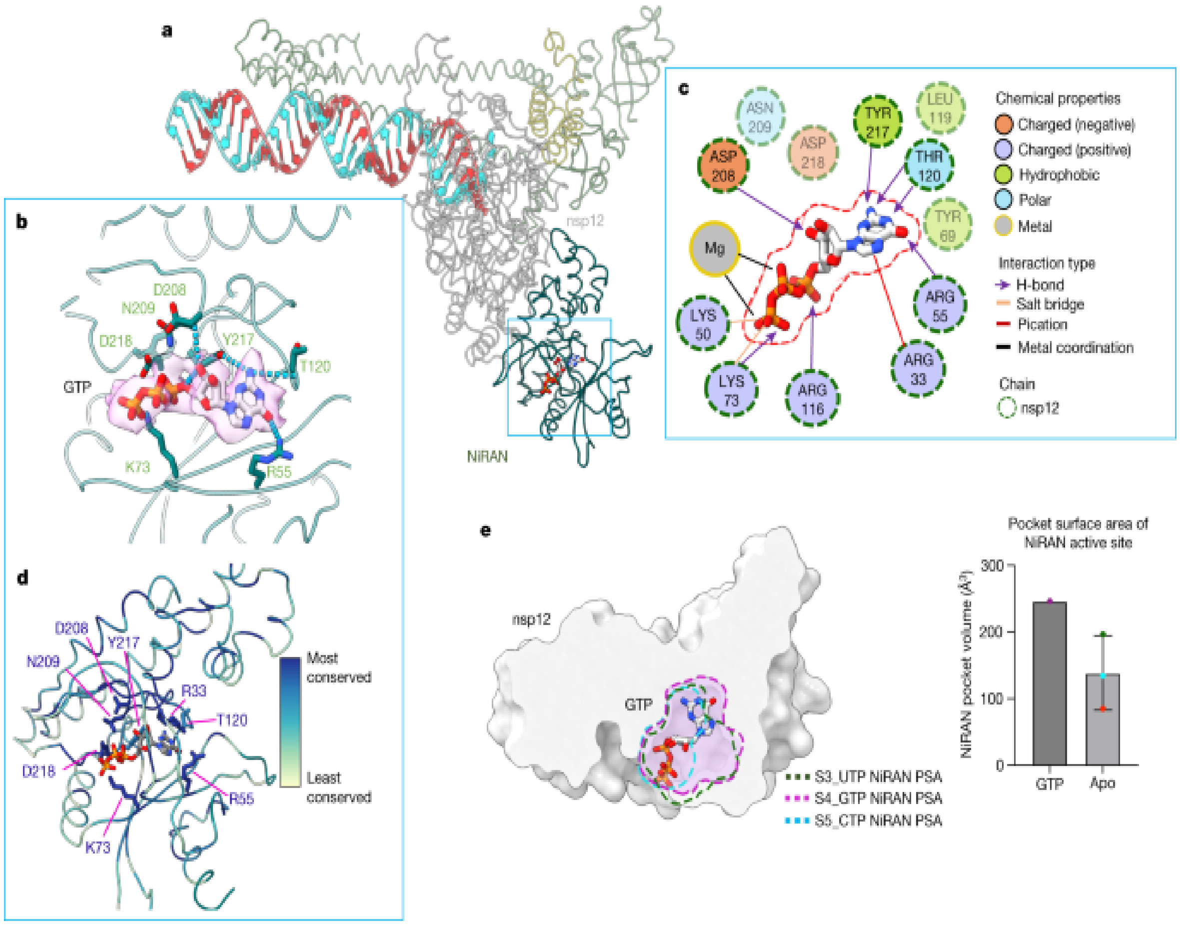
Structural insights into RNA capping

Apart from the RdRp active site, the RTC possesses an additional catalytic domain known as the NiRAN (nidovirus RdRp-associated nucleotidyltransferase), which lies at the N terminal end of the RdRp[5,15,29]. Although the NiRAN domain is essential for viral propagation, its functions

during the viral life cycle remain enigmatic. Recent studies have suggested that the NiRAN functions in the viral RNA capping pathway in conjunction with an additional viral protein, nsp9 (refs. [5,30]). Others have postulated a role for the NiRAN domain in protein-mediated priming of RNA synthesis[31]. In previous structural studies, ADP[15] or GDP[30] have been observed bound in the NiRAN active site in a 'base-out' pose, where the nucleotide base points out of the active-site pocket and makes few, if any, protein contacts. ADP has not been shown to serve as a substrate for the NiRAN enzymatic activity, and the GDP was bound along with an N-terminally modified (and resultingly inactive) nsp9 (ref. [30]), so the relevance of these nucleotide poses is unclear. In another study, the diphosphate form of the GTP analogue AT-9010 is seen bound to the NiRAN in a 'base-in' pose[20], with the phosphates occupying the same positions as the phosphates in the 'base-out' pose but flipped, and with the guanine base bound in a tight pocket internal to the NiRAN (Extended Data Fig. 10).

In our cryo-EM structures of the RTC with each of the natural NTPs and RDV-TP (Supplementary Table 1), only GTP was observed stably and specifically bound in the NiRAN active site. The cryo-EM density supports the presence of GTP (and not GDP) in the NiRAN (Fig. 4). The GTP is bound in the 'base-in' pose similar to the diphosphate form of AT-9010 (ref. [20]), with the base buried in an apparently guanine-specific pocket that extends into the core of the NiRAN fold, enabling it to make key contacts with hydrophilic residues that line the pocket interior. A central element to this recognition is interactions mediated by R55 and Y217, two conserved residues across the α- and β-coronavirus clades, which provide base specificity for guanosine (Fig. 4b,d). The apo-NiRAN pocket is sterically incompatible with the bound guanine base; GTP association is mediated by an induced fit mechanism that involves expansion of the active-site pocket to accommodate the guanosine base (Fig. 4e).

Fig. 4: NiRAN specific recognition of GTP.



a, View of the NiRAN domain of the RTC, bound to GTP, which lies at the amino-terminal end of nsp12. **b**, GTP is selectively recognized in the NiRAN pocket by a series of hydrogen-bonding and electrostatic interactions in which interacting residues are shown as sticks. **c**, 2D schematic illustrating the NiRAN-GTP interactions. **d**, Residues that mediate GTP recognition are shown as sticks and coloured according to their conservation across the α and β coronavirus clades. **e**, Binding of GTP in the NiRAN is mediated through an induced fit that widens the pocket for insertion of the guanine base. Cross-sections (dashed lines) of the pocket surface area (PSA) of the S4_GTP (GTP bound), S3_UTP and S5_CTP (apo-NiRAN) are shown overlaid on a clipped surface view of the nsp12 NiRAN. The NiRAN pocket volume was measured using the Schrodinger Sitemap tool. The mean ± s.d. of the pocket volumes of the S1_RDV-TP, S3_UTP and S5_CTP structures (apo-NiRAN, $n = 3$) is shown. The estimated volume sizes of the datasets are shown as overlaid circles in which

the green, cyan and red circles represent the calculated PSA for the S3, S5 and S1 structures, respectively.

A recent study proposes that the NiRAN mediates two successive steps in the viral RNA capping pathway that initially entails transfer of the nascent RNA 5′ end to the amino terminus of nsp9, forming a covalent RNA-protein intermediate in a process termed RNAylation[6]. A successive polyribonucleotidyltransferase (PRNTase) reaction uses the RNAylated nsp9 as a substrate and transfers the bound RNA to GDP or GTP to produce the cap structure, GpppN-RNA. The PRNTase activity strictly uses GDP or GTP. Our results indicate that the NiRAN uniquely recognizes GTP (but not other NTPs), explaining the base specificity observed for the PRNTase activity.

Discussion

Antiviral nucleotide analogues are highly effective and versatile therapeutics that can be adapted to treat new coronaviral borne diseases as they emerge. Their suitability for repurposing is exemplified by the use of RDV and molnupiravir for the treatment of COVID-19, two analogues that were initially developed as therapeutics against Ebola virus[32] and influenza virus[33], respectively. Repurposing efforts rely on the conserved active site of the RdRp, which has not been observed to mutate as readily as other COVID-19 therapeutic targets such as the Spike protein (Extended Data Fig. 8). Whereas active-site residues are highly conserved across RdRp-encoding viruses, subtle differences can alter the incorporation selectivity of a nucleotide analogue more than 100-fold[4], highlighting the importance of mechanistic investigations of RdRps across the viral realm.

To probe the mechanism of nucleotide recognition by the SARS-CoV-2 RdRp, we determined structures of stalled RdRp complexes containing an RNA primer template, incoming nucleotide substrates or the antiviral RDV-TP, and catalytic metal ions. Structural views of the RTC with each of the respective nucleotides illustrate how the RdRp structurally adapts to proper Watson–Crick base pairing geometry, closing in around the t-RNA/NTP base pair to facilitate catalytic Mg^{2+} coordination and in-line attack of the p-RNA 3′-OH on the NTP α-phosphate.

Through visualizing these structures, we ascertained that the enhanced selectivity for RDV, which primarily manifests biochemically as an effect on K_m, is mediated by the accommodation of its 1′-cyano group in a conserved hydrophilic pocket near the RdRp active site (Fig. 3). This same RdRp active-site pocket harbours an ordered water molecule in the +CTP and +GTP structures (Fig. 3); the bound water is probably present in the +ATP and +UTP complexes as well but not visualized owing to limited resolution (Extended Data Figs. 2 and 3). A mutation that alters the hydrophilicity of this pocket gives rise to RDV resistance[28] and are naturally found in viral families that show reduced sensitivity to RDV[4]. Critically, structural insights into NTP recognition could be exploited in the design of nucleotide analogues as part of structure-based drug design programs that aim to target the coronavirus RdRp. Our studies also shed light on the presence of a guanosine-specific pocket in the essential NiRAN domain, supporting proposals that the NiRAN is involved in the production of capped messenger RNAs as well as further detailing this pocket as a therapeutic target.

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification

SARS-CoV-2 nsp12 was expressed as previously described[34] and purified as follows. Briefly, a pQE-30/pcl-ts ind$^+$ plasmid containing a His_6-small ubiquitin-like modifier (SUMO) SARS-CoV-2 nsp12 and untagged nsp7 and 8 (Addgene no. 160540) was transformed into *Escherichia coli* BL21 cells (Agilent). Cells were grown and protein expression was induced by the addition of 0.2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG), 10 ng ml^{-1} tetracycline and 50 μg ml^{-1} nalidixic acid. Cells were collected and lysed in a French press (Avestin). The lysate was cleared by centrifugation and purified on a HiTrap Heparin HP column (Cytiva). The fractions containing nsp12 were loaded onto a HisTrap HP column (Cytiva) for further purification. Eluted nsp12 was dialysed, cleaved with His_6-Ulp1

SUMO protease and passed through a HisTrap HP column to remove the SUMO protease. Flow-through was collected, concentrated by centrifugal filtration (Amicon) and loaded on a Superdex 200 Hiload 16/600 (Cytiva). Glycerol was added to the purified nsp12, aliquoted, flash-frozen with liquid N_2 and stored at −80 °C.

SARS-CoV-2 nsp7/8 was expressed and purified as described[15]. Briefly, the pCDFDuet-1 plasmid containing His_6 SARS-CoV-2 nsp7/8 (Addgene no. 159092) was transformed into *E. coli* BL21 (DE3). Cells were grown and protein expression was induced by the addition of IPTG. Cells were collected and lysed in a French press (Avestin). The lysate was cleared by centrifugation and purified on a HisTrap HP column (Cytiva). Eluted nsp7/8 was dialysed, cleaved with His_6-Prescission Protease to cleave His_6 tag and then passed through a HisTrap HP column to remove the protease (Cytiva). Flow-through was collected, concentrated by centrifugal filtration (Amicon) and loaded onto a Superdex 75 Hiload 16/600 (Cytiva). Glycerol was added to the purified nsp7/8, aliquoted, flash-frozen with liquid N_2 and stored at −80 °C.

Preparation of SARS-CoV-2 RTC for Cryo-EM

Cryo-EM samples of SARS-CoV-2 RTC were prepared as previously described[15,18]. Briefly, purified nsp12 and nsp7/8 were concentrated, mixed in a 1:3 molar ratio and incubated for 20 min at 22 °C. An annealed RNA scaffold (Horizon Discovery, Ltd) was added to the nsp7/8/12 complex and incubated for 10 min at 30 °C. The sample was buffer exchanged into cryo-EM buffer (20 mM HEPES pH 8.0, 100 mM K-Acetate, 5 mM $MgCl_2$, 2 mM DTT) and further incubated for 20 min at 30 °C. The sample was purified over a Superose 6 Increase 10/300 GL column (Cytiva) in cryo-EM buffer. The peak corresponding to nsp7/8/12/RNA complex was pooled and concentrated by centrifugal filtration (Amicon).

Cryo-EM grid preparation

Before grid freezing, beta-octyl glucoside (β-OG) was added to the sample (0.07% w/v final). The final buffer condition for the cryo-EM sample was

20 mM HEPES pH 8.0, 100 mM K-Acetate, 5 mM $MgCl_2$, 2 mM DTT, 0.07% (w/v) β-OG with 300 µM final of the respective NTP added immediately before freezing. C-flat holey carbon grids (CF-1.2/1.3-4Au, EMS) were glow-discharged for 20 s before the application of 3.5 µl of sample. Using a Vitrobot Mark IV (Thermo Fisher Scientific), grids were blotted and plunge-frozen into liquid ethane with 95% chamber humidity at 4 °C.

Cryo-EM data acquisition and processing

Structural biology software was accessed through the SBGrid consortium[35]. The following pertains to each of the respective datasets:

S1_RDV-TP

Grids were imaged using a 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Images were recorded with Leginon[36] with a pixel size of 1.065 Å per pixel (px) (micrograph dimensions of 5,760 × 4,092 px) over a defocus range of −0.8 to −2.5 µm with a 20 eV energy filter slit. Videos were recorded in 'counting mode' (native K3 camera binning 2) with roughly 25 e^- per px per s in dose-fractionation mode with subframes of 50 ms over a 2.5 s exposure (50 frames) to give a total dose of roughly 54 $e^-/Å^2$. Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2 (ref. [37]). The contrast transfer function (CTF) was estimated for each summed image using the Patch CTF module in cryoSPARC v.3.1.0 (ref. [38]). Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 26,130 motion-corrected images with 8,017,151 particles. Particles were sorted using three rounds of cryoSPARC two-dimensional (2D) classification (*n* = 50, where *n* equals the number of classes), resulting in 2,473,065 curated particles. Initial models (denoted as monomer and dimer) were generated using cryoSPARC ab initio reconstruction (*n* = 3) on a subset of the particles. Particles were further curated using these initial models as three-dimensional (3D) templates for iterative cryoSPARC heterogeneous refinement (*n* = 4), resulting in 314,848

particles in the monomer class (cyan map, Extended Data Fig. 1) and 264,453 particles in the dimer class (pink map, Extended Data Fig. 1). Curated particles in the monomer and dimer classes were re-extracted to a box size of 384 and input to cryoSPARC homogenous and non-uniform refinements[39]. Particles within each class were further processed through two rounds of RELION v.3.1 Bayesian polishing[40]. Polished particles were refined using cryoSPARC local and clobal CTF refinement in combination with cryoSPARC non-uniform refinement, resulting in structures with the following particle counts and nominal resolutions: monomer RTC (251,160 particles; 3.54 Å) and dimer RTC (249,468 particles; 3.87 Å). To facilitate model building of the RTC, particles from the dimer RTC class underwent masked particle subtraction in which both protomers were masked, subtracted from the full map and combined with the original monomer RTC to yield a structure of the RTC at a nominal resolution of 3.38 Å from 613,848 particles (Extended Data Fig. 1e). Local resolution calculations were generated using blocres and blocfilt from the Bsoft package[21] (Extended Data Fig. 1c). The angular distribution of particle orientations (Extended Data Fig. 1b) and directional resolution through the 3DFSC package[41] (Extended Data Fig. 1d) were calculated for the final class.

S2_ATP

Grids were imaged using a C_S (spherical abberation) corrected 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Images were recorded with Leginon[36] with a pixel size of 1.076 Å per px (micrograph dimensions of 5,760 × 4,092 px) over a defocus range of −0.8 to −2.5 μm with a 20 eV energy filter slit. Videos were recorded in counting mode (native K3 camera binning 2) with roughly 25 e^- per px per s in dose-fractionation mode with subframes of 50 ms over a 2.5 s exposure (50 frames) to give a total dose of roughly 52 $e^-/Å^2$. Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2 (ref. 37). The CTF was estimated for each summed image using the patch CTF module in cryoSPARC v.3.1.0 (ref. 38). Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 11,657 motion-corrected images with 5,364,858

particles. Particles were sorted using three rounds of cryoSPARC 2D classification ($n = 50$), resulting in 2,255,856 curated particles. Initial models (denoted as monomer and dimer) were generated using cryoSPARC ab initio reconstruction ($n = 4$) on a subset of the particles. Particles were further curated using these initial models as 3D templates for iterative cryoSPARC heterogeneous refinement ($n = 4$), resulting in 102,529 particles in the monomer class (cyan map, Extended Data Fig. 2) and 318,672 particles in the dimer class (pink map, Extended Data Fig. 2). Curated particles in the monomer and dimer classes were re-extracted to a box size of 384 and input to cryoSPARC homogenous and non-uniform refinements[39]. Particles within each class were further processed through two rounds of RELION v.3.1 Bayesian polishing[40]. Polished particles were refined using cryoSPARC local and clobal CTF refinement in combination with cryoSPARC non-uniform refinement, resulting in structures with the following particle counts and nominal resolutions: monomer RTC (96,868 particles; 3.71 Å) and dimer RTC (299,965 particles; 3.50 Å). To facilitate model building of the RTC, particles from the dimer RTC class underwent masked particle subtraction in which both protomers were masked, subtracted from the full map and combined with the original monomer RTC to yield a structure of the RTC at a nominal resolution of 3.09 Å from 330,442 particles (Extended Data Fig. 2e). This combined RTC class was further refined with masked cryoSPARC local refinement using masks around the RdRp and NiRAN active sites. Locally refined maps were combined into a RTC composite map using PHENIX 'combine focused maps' to aid model building[42]. Local resolution calculations were generated using blocres and blocfilt from the Bsoft package[21] (Extended Data Fig. 2c). The angular distribution of particle orientations (Extended Data Fig. 2b) and directional resolution through the 3DFSC package[41] (Extended Data Fig. 2d) were calculated for the final class.

S3_UTP

Grids were imaged using a 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Images were recorded with Leginon[36] with a pixel size of 1.065 Å per px (micrograph dimensions of 5,760 × 4,092 px) over a defocus range of −0.8 to −2.5 μm with a 20 eV energy filter slit. Videos were recorded in counting mode

(native K3 camera binning 2) with roughly 25 e^- per px per s in dose-fractionation mode with subframes of 50 ms over a 2.5 s exposure (50 frames) to give a total dose of roughly 53 $e^-/Å^2$. Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2 (ref. [37]). The CTF was estimated for each summed image using the patch CTF module in cryoSPARC v.3.1.0 (ref. [38]). Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 30,850 motion-corrected images with 14,149,078 particles. Particles were sorted using three rounds of cryoSPARC 2D classification ($n = 50$), resulting in 3,297,109 curated particles. Initial models (denoted as monomer and dimer) were generated using cryoSPARC ab initio reconstruction ($n = 4$) on a subset of the particles. Particles were further curated using these initial models as 3D templates for iterative cryoSPARC heterogeneous refinement ($n = 6$), resulting in 648,814 particles in the monomer class (cyan map, Extended Data Fig. 3) and 773,911 particles in the dimer class (pink map, Extended Data Fig. 3). Curated particles in the monomer and dimer classes were re-extracted to a box size of 384 and input to cryoSPARC homogenous and non-uniform refinements[39]. Particles within each class were further processed through two rounds of RELION v.3.1 Bayesian polishing[40]. Polished particles were refined using cryoSPARC local and global CTF refinement in combination with cryoSPARC non-uniform refinement, resulting in structures with the following particle counts and nominal resolutions: monomer RTC (614,648 particles; 3.34 Å) and dimer RTC (730,009 particles; 3.42 Å). To facilitate model building of the RTC, particles from the dimer RTC class underwent masked particle subtraction in which both protomers were masked, subtracted from the full map and combined with the original monomer RTC to yield a structure of the RTC at a nominal resolution of 3.13 Å from 719,889 particles (Extended Data Fig. 3e). This combined RTC class was further refined with masked cryoSPARC local refinement using masks around the RdRp and NiRAN active sites. Locally refined maps were combined into a RTC composite map using PHENIX ‘combine focused maps’ to aid model building[42]. Local resolution calculations were generated using blocres and blocfilt from the Bsoft package[21] (Extended Data Fig. 3c). The angular distribution of particle orientations (Extended Data Fig. 3b) and directional resolution through the

3DFSC package[41] (Extended Data Fig. 3d) were calculated for the final class.

S4_GTP

Grids were imaged using a C_S corrected 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Images were recorded with SerialEM[43] with a pixel size of 1.08 Å per px (micrograph dimensions of 5,760 × 4,092 px) over a defocus range of −0.8 to −3.0 μm with a 20 eV energy filter slit. Videos were recorded in counting mode (native K3 camera binning 2) with roughly 25 e^- per px per s in dose-fractionation mode with subframes of 50 ms over a 2.5 s exposure (50 frames) to give a total dose of roughly 51 $e^-/Å^2$. Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2 (ref. [37]). The CTF was estimated for each summed image using the patch CTF module in cryoSPARC v.3.1.0 (ref. [38]). Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 4,527 motion-corrected images with 2,419,929 particles. Particles were sorted using three rounds of cryoSPARC 2D classification ($n = 50$), resulting in 941,507 curated particles. An initial model of the monomer RTC was generated using cryoSPARC ab initio reconstruction ($n = 3$) on a subset of the particles. Particles were further curated using this initial model as a 3D template for iterative cryoSPARC heterogeneous refinement ($n = 6$), resulting in 484,682 particles in the resultant monomer class (cyan map, Extended Data Fig. 4). Curated particles were re-extracted to a box size of 384 and input to cryoSPARC homogenous and non-uniform refinements[39]. Particles were further processed through two rounds of RELION v.3.1 Bayesian polishing[40]. Polished particles were refined using cryoSPARC local and global CTF refinement in combination with cryoSPARC non-uniform refinement, resulting in a structure with the following particle count and nominal resolution: monomer RTC (456,629 particles; 2.68 Å) (Extended Data Fig. 4e). Local resolution calculations were generated using blocres and blocfilt from the Bsoft package[21] (Extended Data Fig. 4c). The angular distribution of particle orientations (Extended Data Fig. 4b) and

directional resolution through the 3DFSC package[41] (Extended Data Fig. 4d) were calculated for the final class.

S5_CTP

Grids were imaged using a C_S corrected 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Images were recorded with SerialEM[43] with a pixel size of 0.515 Å per px (micrograph dimensions of 5,760 × 4,092 px) over a defocus range of −0.8 to −3.0 µm with a 20 eV energy filter slit. Videos were recorded in counting mode (native K3 camera binning 2) with roughly 25 e^- per px per s in dose-fractionation mode with subframes of 50 ms over a 2.5 s exposure (50 frames) to give a total dose of roughly 57 $e^-/Å^2$. Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2 (ref. [37]). The CTF was estimated for each summed image using the Patch CTF module in cryoSPARC v.3.1.0 (ref. [38]). Particles were picked and extracted from the dose-weighted images with box size of 512 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 13,905 motion-corrected images with 1,986,527 particles. Particles were sorted using three rounds of cryoSPARC 2D classification ($n = 50$), resulting in 460,232 curated particles. An initial model (denoted a monomer) was generated using cryoSPARC ab initio reconstruction ($n = 3$) on a subset of the particles. Particles were further curated using these initial models as 3D templates for iterative cryoSPARC heterogeneous refinement ($n = 6$), resulting in 143,110 particles in the monomer class (cyan map, Extended Data Fig. 5) and 93,597 particles in an extracted dimer class (pink map, Extended Data Fig. 5). Curated particles in the monomer and dimer classes were re-extracted to a box size of 768 and input to cryoSPARC homogenous and non-uniform refinements[39]. Particles within each class were further processed through two rounds of RELION v.3.1 Bayesian polishing[40]. Polished particles were refined using cryoSPARC local and global CTF refinement in combination with cryoSPARC non-uniform refinement, resulting in structures with the following particle counts and nominal resolutions: monomer RTC (128,484 particles; 3.09 Å) and dimer RTC (83,555 particles; 3.69 Å). To facilitate model building of the RTC, particles from the dimer RTC class underwent masked particle subtraction in which

both protomers were masked, subtracted from the full map and combined with the original monomer RTC to yield a structure of the RTC at a nominal resolution of 2.67 Å from 171,107 particles (Extended Data Fig. 5e). This combined RTC class was further refined with masked cryoSPARC local refinement using masks around the RdRp and NiRAN active sites. Locally refined maps were combined into a RTC composite map using PHENIX 'combine focused maps' to aid model building[42]. Local resolution calculations were generated using blocres and blocfilt from the Bsoft package[21] (Extended Data Fig. 5c). The angular distribution of particle orientations (Extended Data Fig. 5b) and directional resolution through the 3DFSC package[41] (Extended Data Fig. 5d) are shown for the final class.

Model building and refinement

An initial model of the RTC was derived from Protein Data Bank (PDB) 7RE1 (ref. [18]). The models were manually fit into the cryo-EM density maps using Chimera[44] and rigid-body and real-space refined using PHENIX real-space-refine[42]. For real-space refinement, rigid-body refinement was followed by all-atom and *B* factor refinement with Ramachandran and secondary structure restraints. Models were inspected and modified in Coot v.0.9.5 (ref. [45]) and the refinement process was repeated iteratively.

nMS analysis

The RTC samples were initially reconstituted as described above in the following buffer: 20 mM HEPES pH 8.0, 80 mM K-Acetate, 5 mM $MgCl_2$ and 2 mM DTT. For the NTP pre-incorporation experiments, the RTCs with RNA scaffolds containing either 3′-oxy p-RNA (RTC) or 3′-deoxy p-RNA (RTC*) were incubated with 300 µM NTP (ATP, RDV-TP or GTP) on ice for 2 min before buffer exchange. For the NDP incubation experiments, the RTC samples reconstituted with the respective RNA scaffolds were incubated with 2 mM nucleotide (ADP, RDV-DP, GDP or ATP) on ice for 2 min before buffer exchange. After nucleotide incubation, all samples were immediately buffer exchanged into an nMS-compatible solution (150 mM ammonium acetate, pH 7.5, 0.01% Tween-20) with a 40 kDa molecular weight cut-off (Thermo Fisher Scientific). For nMS analysis, a 2–3 µl

aliquot of the buffer-exchanged sample was loaded into a gold-coated quartz capillary tip prepared in-house and then electrosprayed into an Exactive Plus with an extended mass range instrument (Thermo Fisher Scientific) with a static direct infusion nanospray source[46]. The MS parameters used include: spray voltage, 1.22 kV; capillary temperature, 125–150 °C; in-source dissociation, 0–10 V; S-lens radio-frequency level, 200; resolving power, 8,750 or 17,500 at an *m/z* of 200; AGC target, 1×10^6; maximum injection time, 200 ms; number of microscans was five; injection flatapole, 8 V; interflatapole, 7 V; bent flatapole, 4 V; high energy collision dissociation (HCD), 200 V; ultrahigh vacuum pressure, $5.8–6.1 \times 10^{-10}$ mbar and total number of scans was at least 100. Mass calibration in positive extended mass range mode was performed using caesium iodide. For data processing, the acquired MS spectra were visualized using Thermo Xcalibur Qual Browser (v.4.2.47). Deconvolution was performed either manually or using the software UniDec v.4.2.0 (refs. [47,48]). The following parameters were used for data processing with UniDec: background subtraction (if applied) was a subtract curve of ten; smooth charge state distribution was enabled and peak shape function was Gaussian. Mass accuracies were calculated as the percentage difference between the measured and expected masses relative to the expected mass. The observed mass accuracies (calculated as the percentage difference between the measured and expected masses relative to the expected mass) ranged from 0.005 to 0.06%.

The expected masses for the component proteins are nsp7, 9,137 Da; nsp8 (N-terminal Met lost), 21,881 Da and nsp12 (has two Zn^{2+} ions coordinated with six deprotonated cysteine residues), 106,785 Da (ref. [15]). The RNA scaffolds were also analysed separately, and their sequences were verified by mass measurements using nMS.

In vitro primer elongation assays

Assays were performed using reconstituted template-primer RNA scaffolds (Supplementary Table 1) (Horizon Discover Ltd/Dharmacon) annealed in 10 mM HEPES pH 8.0, 50 mM KCl and 2 mM $MgCl_2$. Reactions (20 µl) containing 100 nM RNA scaffold, 0.75 µM nsp12, 2 µM nsp7/8 and NTPs 300 µM (if present as natural nucleotides, NDPs or α-β analogues) and 1 µl of α-^{32}P-GTP (Perkin-Elmer) were incubated at 30 °C for 10 min before

addition of a 2× stop solution (Invitrogen-Gel Loading buffer II). Assay buffer was 100 mM K-acetate, 20 mM HEPES pH 8.0, 5 mM $MgCl_2$ and 2 mM beta-mercaptoethanol in which the $MgCl_2$ was substituted for $CaCl_2$ when monitoring the effects of alternative metal ions. Products of the elongation reactions were separated on 10% acrylamide-8M urea denaturing gels and analysed by phosphorimaging.

Substrate selectivity assays

The selectivity of nucleotide incorporation by SARS-CoV-2 wild-type and mutant RdRps was determined, and materials and reagents acquired as previously described[4,49]. The concentration for all wild-type and mutant nsp12 enzymes per reaction was 150 nM, 750 nM for nsp7 and 1.5 µM for nsp8. Nsp12, nsp7 and nsp8 were assembled in reaction. Reactions were incubated for 10 min to allow for single-nucleotide incorporation. Michaelis–Menten kinetics analyses were applied as described. Nucleotide incorporation efficiencies (catalytic efficiencies) were calculated by taking the ratio between V_{max} and K_m. Selectivity values were determined from the ratio between ATP incorporation efficiency to that of RDV-TP. Finally, the discrimination index was calculated by taking the ratios of the mutant RdRps selectivity values to that of the wild-type one[4,49].

Quantification and statistical analysis

The local resolution of the cryo-EM maps (Extended Data Figs. 1–5) was estimated using blocres[21] with the following parameters: box size 15 and sampling with respective map pixel size. Directional 3DFSCs (Extended Data Figs. 1–5) were calculated using 3DFSC (ref. [41]). The quantification and statistical analyses for model refinement and validation were generated using MolProbity[50] and PHENIX[42].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All unique/stable reagents generated in this study are available without restriction from the lead contact, S.A.D. (darst@rockefeller.edu). The cryo-EM density maps and atomic coordinates have been deposited in the EMDataBank and PDB as follows: S1_RDV-TP (EMD-26639, PDB 7UO4), S2_ATP (EMD-26641, PDB 7UO7), S3_UTP (EMD-26642, PDB 7UO9), S4_GTP (EMD-26645, PDB 7UOB) and S5_CTP (EMD-26646, PDB 7UOE). The model used for initial building (PDB 7RE1) is available on the PDB. The atomic models used for initial model building and analysis are available from the Protein Data Bank under the accession codes 7RE1, 6XEZ, 7CYQ and 7ED5.

References

1. Malone, B., Urakova, N., Snijder, E. J. & Campbell, E. A. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.* **23**, 21–39 (2022).

2. Beigel, J. H. et al. Remdesivir for the treatment of COVID-19—final report. *N. Engl. J. Med.* **383**, 1813–1826 (2020).

3. Dangerfield, T. L., Huang, N. Z. & Johnson, K. A. Remdesivir is effective in combating COVID-19 because it is a better substrate than ATP for the viral RNA-dependent RNA polymerase. *iScience* **23**, 101849 (2020).

4. Gordon, C. J. et al. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* **295**, 6785–6797 (2020).

5. Slanina, H. et al. Coronavirus replication–transcription complex: vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. *Proc. Natl Acad. Sci. USA* **118**, e2022310118 (2021).

6. Park, G. J. et al. The mechanism of RNA capping by SARS-CoV-2. *Nature* **609**, 793–800 (2022).

7. Jayk Bernal, A. et al. Molnupiravir for oral treatment of COVID-19 in nonhospitalized patients. *N. Engl. J. Med.* **386**, 509–520 (2021).

8. Bruenn, J. A. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* **31**, 1821–1829 (2003).

9. Bravo, J. P. K., Dangerfield, T. L., Taylor, D. W. & Johnson, K. A. Remdesivir is a delayed translocation inhibitor of SARS CoV-2 replication. *Mol. Cell* https://doi.org/10.1016/j.molcel.2021.01.035 (2021).

10. Seifert, M. et al. Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *eLife* **10**, e70968 (2021).

11. Steitz, T. A. DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395–17398 (1999).

12. Appleby, T. C. et al. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347**, 771 (2015).

13. Shu, B. & Gong, P. Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *Proc. Natl Acad. Sci. USA* **113**, E4005–E4014 (2016).

14. Zamyatkin, D. F. et al. Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *J. Biol. Chem.* **283**, 7705–7712 (2008).

15. Chen, J. et al. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* https://doi.org/10.1016/j.cell.2020.07.033 (2020).

16. Hillen, H. S. et al. Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154–156 (2020).

17. Wang, Q. et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* **182**, 417–428.e413 (2020).

18. Chen, J. et al. Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication–transcription complex. *Nat. Struct. Mol. Biol.* **29**, 250–260 (2022).

19. Naydenova, K. et al. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc. Natl Acad. Sci. USA* **118**, e2021946118 (2021).

20. Shannon, A. et al. A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. *Nat. Commun.* **13**, 621 (2022).

21. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).

22. Salgado, P. S. et al. The structural basis for RNA specificity and Ca^{2+} inhibition of an RNA-dependent RNA polymerase. *Structure* **12**, 307–316 (2004).

23. Temiakov, D. et al. Structural basis for substrate selection by T7 RNA polymerase. *Cell* **116**, 381–391 (2004).

24. Wang, M. et al. SARS-CoV-2 RdRp uses NDPsas a substrate and is able to incorporate NHC into RNA by using diphosphate form molnupiravir. *Int. J. Biol. Macromol.* **226**, 946–955 (2023).

25. Yang, W., Lee, J. Y. & Nowotny, M. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol. Cell* **22**, 5–13 (2006).

26. Ben Ouirane, K., Boulard, Y. & Bressanelli, S. The hepatitis C virus RNA-dependent RNA polymerase directs incoming nucleotides to its active site through magnesium-dependent dynamics within its F motif. *J. Biol. Chem.* **294**, 7573–7587 (2019).

27. Malone, B. et al. Structural basis for backtracking by the SARS-CoV-2 replication–transcription complex. *Proc. Natl Acad. Sci. USA* **118**, e2102516118 (2021).

28. Stevens, L. J. et al. Mutations in the SARS-CoV-2 RNA dependent RNA polymerase confer resistance to remdesivir by distinct mechanisms. *Sci. Transl. Med.* https://doi.org/10.1126/scitranslmed.abo0718 (2022).

29. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 2342 (2019).

30. Yan, L. et al. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell* **184**, 184–193.e110 (2021).

31. Shannon, A. et al. Protein-primed RNA synthesis in SARS-CoVs and structural basis for inhibition by AT-527. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.23.436564 (2021).

32. Mulangu, S. et al. A randomized, controlled trial of Ebola virus disease therapeutics. *N. Engl. J. Med.* **381**, 2293–2303 (2019).

33. Painter, G. R., Natchus, M. G., Cohen, O., Holman, W. & Painter, W. P. Developing a direct acting, orally available antiviral agent in a pandemic: the evolution of molnupiravir as a potential treatment for COVID-19. *Curr. Opin. Virol.* **50**, 17–22 (2021).

34. Dangerfield, T. L., Huang, N. Z. & Johnson, K. A. Expression and purification of tag-free SARS-CoV-2 RNA-dependent RNA polymerase in *Escherichia coli*. *STAR Protoc.* **2**, 100357 (2021).

35. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).

36. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).

37. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

38. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

39. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).

40. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

41. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

42. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).

43. Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).

44. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

46. Olinares, P. D. B. & Chait, B. T. Native mass spectrometry analysis of affinity-captured endogenous yeast RNA exosome complexes. *Methods Mol. Biol.* **2062**, 357–382 (2020).

47. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal.*

Chem. **87**, 4370–4376 (2015).

48. Reid, D. J. et al. MetaUniDec: high-throughput deconvolution of native mass spectra. *J. Am. Soc. Mass. Spectrom.* **30**, 118–127 (2019).

49. Gordon, C. J., Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Götte, M. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J. Biol. Chem.* **295**, 4773–4779 (2020).

50. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

Acknowledgements

We thank A. Mueller and R. Saecker for helpful discussions. Some of the work reported here was conducted at the Simons Electron Microscopy Center and the National Resource for Automated Molecular Microscopy and National Center for CryoEM Access and Training located at the NYSBC, supported by grants from the National Insitutes of Health (NIH) National Institute of General Medical Sciences (grant no. P41 GM103310), NYSTAR, the Simons Foundation (grant no. SF349247), the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy programme (grant no. U24 GM129539) and NY State Assembly Majority. This work was supported by NIH grant nos. P41 GM109824 and P41 GM103314 to B.T.C., and NIH grant no. R01 AI161278 and Gilead Sciences (to E.A.C. and S.A.D.). M.G. received funding from the Canadian Institutes of Health Research (CIHR), grant no. 170343, and the Alberta Ministry of Jobs, Economy and Northern Development through the Major Innovation Fund on Antimicrobial Resistance (AMR) - One Health Consortium.

Author information

Author notes

1. James Chen

   Present address: Department of Cell Biology, New York University School of Medicine, New York, NY, USA

Authors and Affiliations

1. Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY, USA

   Brandon F. Malone, James Chen, Elizabeth A. Campbell & Seth A. Darst

2. Gilead Sciences, Inc., Foster City, CA, USA

   Jason K. Perry, Todd C. Appleby, Joy Y. Feng & John P. Bilello

3. Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, USA

   Paul Dominic B. Olinares & Brian T. Chait

4. Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta, Canada

   Hery W. Lee & Matthias Götte

5. The Evelyn Gruss Lipper Cryo-Electron Microscopy Resource Center, The Rockefeller University, New York, NY, USA

   Honkit Ng, Johanna Sotiris & Mark Ebrahim

6. National Center for Cryo-EM Access and Training, Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA

   Eugene Y. D. Chua, Joshua H. Mendez & Ed T. Eng

7. Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

Robert Landick

Contributions

B.F.M., J.C., J.K.P., T.C.A., J.Y.F., J.P.B., E.A.C. and S.A.D. conceived and designed this study. B.F.M., H.W.L. and J.C. performed cloning, protein purification and biochemistry. P.D.B.O. conducted mass spectrometry experiments. B.F.M. prepared cryo-EM specimens. Cryo-EM data were collected by B.F.M., E.Y.D.C., J.H.M., E.T.E., M.E., J.S. and H.N. B.F.M. processed all cryo-EM data. B.F.M., J.K.P., E.A.C. and S.A.D. built and analysed atomic models. E.A.C., R.L., B.T.C., M.G. and S.A.D. supervised and acquired financial support. B.F.M. wrote the first draft of the manuscript; all authors contributed to the final version.

Corresponding authors

Correspondence to Elizabeth A. Campbell or Seth A. Darst.

Ethics declarations

Competing interests

E.A.C. and S.A.D. received funding from Gilead Sciences, Inc. in support of this study. J.K.P., T.C.A., J.Y.F. and J.P.B. are Gilead employees. M.G. received funding from Gilead Sciences Inc. for studies on the mechanism of action of remdesivir. All other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Dong Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are

available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Cryo-EM processing pipeline and analysis for S1_RDV-TP dataset.

(**a**) Cryo-EM processing pipeline for the S1_RDV-TP dataset. (**b**) Angular distribution plot for the S1_RDV-TP dataset, calculated in cryoSPARC. Scale depicts number of particles assigned to a specific angular bin. (**c**) Nominal 3.38 Å resolution cryo-EM reconstruction filtered by local resolution and colored according to fitted model chain. Right panel is clipped to reveal RTC active site. (**d**) Directional 3D FSC for S1_RDV-TP, determined with 3DFSC. (**e**) Gold-standard FSC plot for the S1_RDV-TP dataset, calculated by comparing two half maps from cryoSPARC. The blue dotted line represents the 0.143 FSC cutoff.

Extended Data Fig. 2 Cryo-EM processing pipeline and analysis for S2_ATP dataset.

(**a**) Cryo-EM processing pipeline for the S2_ATP dataset. (**b**) Angular distribution plot for the S2_ATP dataset, calculated in cryoSPARC. Scale depicts number of particles assigned to a specific angular bin. (**c**) Nominal 3.09 Å resolution cryo-EM reconstruction filtered by local resolution and colored according to fitted model chain. Right panel is clipped to reveal RTC active site. (**d**) Directional 3D FSC for S2_ATP, determined with 3DFSC. (**e**) Gold-standard FSC plot for the S2_ATP dataset, calculated by comparing two half maps from cryoSPARC. The blue dotted line represents the 0.143 FSC cutoff.

Extended Data Fig. 3 Cryo-EM processing pipeline and analysis for S3_UTP dataset.

(**a**) Cryo-EM processing pipeline for the S3_UTP dataset. (**b**) Angular distribution plot for the S3_UTP dataset, calculated in cryoSPARC. Scale depicts number of particles assigned to a specific angular bin. (**c**) Nominal 3.13 Å resolution cryo-EM reconstruction filtered by local resolution and colored according to fitted model chain. Right panel is clipped to reveal RTC active site. (**d**) Directional 3D FSC for S3_UTP, determined with 3DFSC. (**e**) Gold-standard FSC plot for the S3_UTP dataset, calculated by comparing two half maps from cryoSPARC. The blue dotted line represents the 0.143 FSC cutoff.

Extended Data Fig. 4 Cryo-EM processing pipeline and analysis for S4_GTP dataset.

(**a**) Cryo-EM processing pipeline for the S4_GTP dataset. (**b**) Angular distribution plot for the S4_GTP dataset, calculated in cryoSPARC. Scale depicts number of particles assigned to a specific angular bin. (**c**) Nominal 2.68 Å resolution cryo-EM reconstruction filtered by local resolution and colored according to fitted model chain. Right panel is clipped to reveal RTC active site. (**d**) Directional 3D FSC for S4_GTP, determined with 3DFSC. (**e**) Gold-standard FSC plot for the S4_GTP dataset, calculated by comparing two half maps from cryoSPARC. The black dotted line represents the 0.143 FSC cutoff.

Extended Data Fig. 5 Cryo-EM processing pipeline and analysis for S5_CTP dataset.

(**a**) Cryo-EM processing pipeline for the S5_CTP dataset. (**b**) Angular distribution plot for the S5_CTP dataset, calculated in cryoSPARC. Scale depicts number of particles assigned to a specific angular bin. (**c**) Nominal 2.67 Å resolution cryo-EM reconstruction filtered by local resolution and colored according to fitted model chain. Right panel is clipped to reveal RTC active site. (**d**) Directional 3D FSC for S5_CTP, determined with 3DFSC. (**e**) Gold-standard FSC plot for the S5_CTP dataset, calculated by

comparing two half maps from cryoSPARC. The black dotted line represents the 0.143 FSC cutoff.

Extended Data Fig. 6 Biochemical assays probing stabilization of pre-incorporation complex.

(**a**) SDS-PAGE of purified SARS-CoV-2 nsp7/8 & nsp12. The experiment was performed once. (**b**) Size exclusion chromatography for the purified RTC complex, composed of nsp7/8$_2$/12 bound to the reconstituted product/template-RNA scaffold (chromatogram trace using the S4 scaffold is depicted). (**c**) SDS-PAGE of assembled RTC complex following size exclusion. Gel (n = 1) illustrates RTC assembly on the S4 scaffold. (**d**) S4 RNA scaffold utilized for the S4_GTP structure and incorporation assays. (**e**) Gel-based primer elongation assay in presence of physiological metal, Mg^{2+}, and non-physiological metal, Ca^{2+}, to investigate use of Ca^{2+} for stabilization of the pre-incorporation complex (gel depicts n = 2 with n = 3 experiments performed). Products are visualized alongside Decade RNA ladder (Invitrogen) showing separation in nucleotides (nts). (**f**) S3 RNA scaffold used for the S3_UTP structure and incorporation assays. (**g**) Gel-based primer elongation assay in presence of UTP, UMPNPP and 3′deoxy UTP (gel depicts n = 2 with n = 3 experiments performed). Products are visualized alongside Decade RNA ladder (Invitrogen) showing separation in nucleotides (nts). (**h**) Native mass spectrometry analysis of RNA extension in presence of ADP, RDV-DP & ATP using the S1/S2 RNA scaffold. (**i**) Native mass spectrometry analysis of RNA extension in presence of GDP using the S4 RNA scaffold.

Extended Data Fig. 7 Motif F residue interactions with the incoming NTPs.

(**a**–**e**) Analysis of the active sites of structures (S1–S5) illustrate that K545 and R555 interact differentially with each of the incoming nucleotides. H-bonds are depicted as dotted-cyan lines. (**f**,**g**) 2D schematics illustrating the set of interactions between RDV-TP (**f**) and ATP (**g**) and the active site.

Extended Data Fig. 8 Conservation of the CoV active site across clades and SARS-CoV-2 strains.

(**a**) Zoom-in on the active site of S4_GTP, highlighting residues (sticks) which interact with the bound GTP/2xMg in which ribbon/sticks are colored according to the amino acid conservation across a representative list of viruses found in the α & β coronavirus clades (Supplementary Table 3). (**b**) Bar plot showing the frequency of occurrence of the wild-type amino acid (reference strain Wuhan/Hu-1/2019) as well as their summed mutations according to the GISAID database, as of April 2022, for the nsp12 active site residues and the nsp12 residue P323. (**c**) Bar plot showing the frequency of occurrence of the wild-type amino acid (reference strain Wuhan/Hu-1/2019) as well as their summed mutations according to the GISAID database, as of April 2022, for Spike (S) residues found in the ACE2 binding region.

Extended Data Fig. 9 Biochemical analysis of the role of residue S759 in RDV-TP recognition.

(**a**) RNA template sequence used to determine the efficiency of adenosine triphosphate (ATP) or RDV-TP incorporation at position 6 (i). G indicates incorporation of [α32P] guanosine triphosphate (GTP) at position 5 (red). (**b**) Migration patterns of the products of ATP or RDV-TP incorporation reactions with wild-type and S759A SARS-CoV-2 RdRp complexes are shown. Main products emerge at position 6. The 5’- 32P-labeled 4-nucleotide primer (4) is used as a size marker (m), illustrating sizing in nucleotides (nts) (**c**) Graphical representation of ATP or RDV-TP single-nucleotide incorporation during RNA synthesis as a function of their respective concentrations shown in (**b**). Best-fit lines illustrate fitting of the data points to Michaelis-Menten kinetics function using GraphPad Prism 7.0. Data shows mean values with error bars illustrating the SDs of the data. All data represent at least three independent experiments (n = 4 for wild-type and n = 3 for S759A SARS-CoV-2 RdRp complexes). (**d**) V_{max} refers to the maximum velocity required to convert substrates into products in a substrate-saturated system, reported as a product fraction of incorporated nucleotide. K_m is a parameter indicating the concentration of substrate at

one-half V_{max}, in µM. The standard error of the linear fit is denoted (±). The V_{max}/K_m ratio reflects catalytic efficiency and is used here to determine the selectivity of RDV-TP by taking the ratio between the catalytic efficiency of ATP and of RDV-TP. The discrimination index is defined as the ratio between the mutant RdRp's selectivity values in relation to that of the wild-type.

Extended Data Fig. 10 Comparison of NiRAN binding poses.

(**a**) Models (PDBs 6XEZ & 7CYQ) of the nsp12 NiRAN with a bound base in the 'Base-out' pose. (**b**) Models (PDBs 7ED5 & 7UOB) of the nsp12 NiRAN with a bound base in the 'Base-in' pose. (**c**) Surface representation detailing the NiRAN active site pocket bound to a nucleotide in the 'base-out' pose (PDB 6XEZ) aligned with a structure with a bound nucleotide in the 'base-in' pose (PDB 7ED5).

Supplementary information

Supplementary Information

Supplementary Figs. 1–3, Tables 1–3 and legend for Video 1.

Reporting Summary.

Peer Review File.

Supplementary Video 1

Nucleotide binding induces active closure, as illustrated through a comparison of an apo RdRp complex (PDB 7RE1) with PDB 7UOB (+GTP structure).

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

About this article

Cite this article

Malone, B.F., Perry, J.K., Olinares, P.D.B. *et al.* Structural basis for substrate selection by the SARS-CoV-2 replicase. *Nature* **614**, 781–787 (2023). https://doi.org/10.1038/s41586-022-05664-3

- Received: 09 May 2022
- Accepted: 15 December 2022
- Published: 01 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-022-05664-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05664-3

Amendments & Corrections

- **Author Correction: Long-primed germinal centres with enduring affinity maturation and clonal migration** [30 Jan 2023]
  -
- **Author Correction: SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry** [06 Feb 2023]
  -
- **Author Correction: Senescence atlas reveals an aged-like inflamed niche that blunts muscle regeneration** [06 Feb 2023]
  -
- **Publisher Correction: Future temperature extremes threaten land vertebrates** [06 Feb 2023]
  -

- Author Correction
- Published: 30 January 2023

Author Correction: Long-primed germinal centres with enduring affinity maturation and clonal migration

- Jeong Hyun Lee,
- Henry J. Sutton,
- Christopher A. Cottrell,
- Ivy Phung,
- Gabriel Ozorowski,
- Leigh M. Sewall,
- Rebecca Nedellec,
- Catherine Nakao,
- Murillo Silva,
- Sara T. Richey,
- Jonathan L. Torres,
- Wen-Hsin Lee,
- Erik Georgeson,
- Michael Kubitz,
- Sam Hodges,
- Tina-Marie Mullen,
- Yumiko Adachi,
- Kimberly M. Cirelli,
- Amitinder Kaur,
- Carolina Allers,
- Marissa Fahlberg,
- Brooke F. Grasperge,

- Jason P. Dufour,
- Faith Schiro,
- Pyone P. Aye,
- Oleksandr Kalyuzhniy,
- Alessia Liguori,
- Diane G. Carnathan,
- Guido Silvestri,
- Xiaoying Shen,
- David C. Montefiori,
- Ronald S. Veazey,
- Andrew B. Ward,
- Lars Hangartner,
- Dennis R. Burton,
- Darrell J. Irvine,
- William R. Schief &
- …
- Shane Crotty

Show authors

Nature **volume 614**, page E43 (2023)

- 559 Accesses
- 11 Altmetric
- Metrics details

The Original Article was published on 21 September 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05216-9
Published online 21 September 2022

In the version of this article initially published, the surname of Jonathan L. Torres was misspelled as “Torre.” In addition, the *x* axis of panel d in Figure 1, now reading “MD39 SA-PE,” was incorrectly labelled “MD39 SA-

PEEnv.” Furthermore, the lettered panel labels in Extended Data Figure 9 did not match the corresponding figure legend. The errors have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Jeong Hyun Lee, Henry J. Sutton

Authors and Affiliations

1. Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La Jolla, CA, USA

   Jeong Hyun Lee, Henry J. Sutton, Ivy Phung, Catherine Nakao, Kimberly M. Cirelli & Shane Crotty

2. Consortium for HIV/AIDS Vaccine Development, The Scripps Research Institute, La Jolla, CA, USA

   Jeong Hyun Lee, Henry J. Sutton, Christopher A. Cottrell, Ivy Phung, Gabriel Ozorowski, Murillo Silva, Erik Georgeson, Michael Kubitz, Sam Hodges, Tina-Marie Mullen, Yumiko Adachi, Kimberly M. Cirelli, Oleksandr Kalyuzhniy, Alessia Liguori, Diane G. Carnathan, Guido Silvestri, Andrew B. Ward, Lars Hangartner, Dennis R. Burton, Darrell J. Irvine, William R. Schief & Shane Crotty

3. IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA, USA

   Jeong Hyun Lee, Christopher A. Cottrell, Gabriel Ozorowski, Erik Georgeson, Michael Kubitz, Sam Hodges, Tina-Marie Mullen, Yumiko Adachi, Oleksandr Kalyuzhniy, Alessia Liguori, Andrew B. Ward, Dennis R. Burton & William R. Schief

4. Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA

Christopher A. Cottrell, Rebecca Nedellec, Erik Georgeson, Michael Kubitz, Sam Hodges, Tina-Marie Mullen, Yumiko Adachi, Oleksandr Kalyuzhniy, Alessia Liguori, Lars Hangartner, Dennis R. Burton & William R. Schief

5. Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California, San Diego, La Jolla, CA, USA

Ivy Phung & Shane Crotty

6. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA

Gabriel Ozorowski, Leigh M. Sewall, Sara T. Richey, Jonathan L. Torres, Wen-Hsin Lee & Andrew B. Ward

7. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Murillo Silva & Darrell J. Irvine

8. Tulane National Primate Research Center, Tulane School of Medicine, Covington, LA, USA

Amitinder Kaur, Carolina Allers, Marissa Fahlberg, Brooke F. Grasperge, Jason P. Dufour, Faith Schiro, Pyone P. Aye & Ronald S. Veazey

9. Emory National Primate Research Center and Emory Vaccine Center, Emory University School of Medicine, Atlanta, GA, USA

Diane G. Carnathan & Guido Silvestri

10. Department of Surgery, Laboratory for AIDS Vaccine Research & Development, Duke University Medical Center, Duke University, Durham, NC, USA

Xiaoying Shen & David C. Montefiori

11. Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA

    Dennis R. Burton, Darrell J. Irvine & William R. Schief

12. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

    Darrell J. Irvine

13. Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

    Darrell J. Irvine

14. Howard Hughes Medical Institute, Chevy Chase, MD, USA

    Darrell J. Irvine

Corresponding author

Correspondence to Shane Crotty.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Lee, J.H., Sutton, H.J., Cottrell, C.A. *et al.* Author Correction: Long-primed germinal centres with enduring affinity maturation and clonal migration. *Nature* **614**, E43 (2023). https://doi.org/10.1038/s41586-023-05741-1

- Published: 30 January 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05741-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05741-1

- Author Correction
- Published: 06 February 2023

Author Correction: SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry

- John Kee,
- Samuel Thudium,
- David M. Renner,
- Karl Glastad,
- Katherine Palozola,
- Zhen Zhang,
- Yize Li,
- Yemin Lan,
- Joseph Cesare,
- Andrey Poleshko,
- Anna A. Kiseleva,
- Rachel Truitt,
- Fabian L. Cardenas-Diaz,
- Xianwen Zhang,
- Xuping Xie,
- Darrell N. Kotton,
- Konstantinos D. Alysandratos,
- Jonathan A. Epstein,
- Pei-Yong Shi,
- Wenli Yang,
- Edward Morrisey,
- Benjamin A. Garcia,
- Shelley L. Berger,
- Susan R. Weiss &
- …

- Erica Korb

Show authors

Nature **volume 614**, page E44 (2023)

- 527 Accesses

- 1 Altmetric

- Metrics details

The Original Article was published on 05 October 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05282-z
Published online 5 October 2022

In the version of this article originally published, the first name of Jonathan A. Epstein was misspelled "Johnathan". The error has been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Samuel Thudium, David M. Renner, Karl Glastad

Authors and Affiliations

1. Department of Genetics at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   John Kee, Samuel Thudium, Katherine Palozola & Erica Korb

2. Epigenetics Institute at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   John Kee, Samuel Thudium, Karl Glastad, Katherine Palozola, Zhen Zhang, Yemin Lan, Joseph Cesare, Benjamin A. Garcia, Shelley L. Berger & Erica Korb

3. Department of Microbiology at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   David M. Renner, Yize Li & Susan R. Weiss

4. Penn Center for Research on Coronaviruses and Other Emerging Pathogens at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   David M. Renner, Yize Li & Susan R. Weiss

5. Department of Cell and Developmental Biology at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Karl Glastad, Zhen Zhang, Andrey Poleshko, Anna A. Kiseleva, Jonathan A. Epstein & Shelley L. Berger

6. Department of Biochemistry and Biophysics at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Joseph Cesare & Benjamin A. Garcia

7. Penn Cardiovascular Institute at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Anna A. Kiseleva & Jonathan A. Epstein

8. Institute for Regenerative Medicine at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Anna A. Kiseleva & Jonathan A. Epstein

9. Department of Medicine at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Rachel Truitt, Fabian L. Cardenas-Diaz, Jonathan A. Epstein, Wenli Yang & Edward Morrisey

10. Penn-CHOP Lung Biology Institute at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

    Fabian L. Cardenas-Diaz & Edward Morrisey

11. Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA

    Xianwen Zhang, Xuping Xie & Pei-Yong Shi

12. Center for Regenerative Medicine, Boston University and Boston Medical Center, Boston, MA, USA

    Darrell N. Kotton & Konstantinos D. Alysandratos

13. The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA, USA

    Darrell N. Kotton & Konstantinos D. Alysandratos

14. Department of Biology at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

    Shelley L. Berger

Corresponding author

Correspondence to Erica Korb.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kee, J., Thudium, S., Renner, D.M. *et al.* Author Correction: SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry. *Nature* **614**, E44 (2023). https://doi.org/10.1038/s41586-023-05764-8

- Published: 06 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05764-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05764-8

- Author Correction
- Open Access
- Published: 06 February 2023

Author Correction: Senescence atlas reveals an aged-like inflamed niche that blunts muscle regeneration

- Victoria Moiseeva,
- Andrés Cisneros,
- Valentina Sica,
- Oleg Deryagin,
- Yiwei Lai,
- Sascha Jung,
- Eva Andrés,
- Juan An,
- Jessica Segalés,
- Laura Ortet,
- Vera Lukesova,
- Giacomo Volpe,
- Alberto Benguria,
- Ana Dopazo,
- Salvador Aznar Benitah,
- Yasuteru Urano,
- Antonio del Sol,
- Miguel A. Esteban,
- Yasuyuki Ohkawa,
- Antonio L. Serrano,
- Eusebio Perdiguero &

- …
- Pura Muñoz-Cánoves

Show authors

Nature **volume 614**, page E45 (2023)

- 726 Accesses
- 2 Altmetric
- Metrics details

The Original Article was published on 21 December 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05535-x
Published online 21 December 2022

In the version of this article initially published, the middle name and surname of Salvador Aznar Benitah were wrongly hyphenated, reading as "Aznar-Benitah". The error has been corrected in the HTML and PDF versions of the article.

Author information

Authors and Affiliations

1. Department of Medicine and Life Sciences, Pompeu Fabra University, Barcelona, Spain

   Victoria Moiseeva, Andrés Cisneros, Valentina Sica, Oleg Deryagin, Eva Andrés, Jessica Segalés, Laura Ortet, Vera Lukesova, Antonio L. Serrano, Eusebio Perdiguero & Pura Muñoz-Cánoves

2. CIBERNED, Barcelona, Spain

Victoria Moiseeva, Andrés Cisneros, Valentina Sica, Oleg Deryagin, Eva Andrés, Jessica Segalés, Laura Ortet, Vera Lukesova, Antonio L. Serrano, Eusebio Perdiguero & Pura Muñoz-Cánoves

3. Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

Yiwei Lai, Juan An, Giacomo Volpe & Miguel A. Esteban

4. Key Laboratory of Regenerative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

Yiwei Lai, Juan An, Giacomo Volpe & Miguel A. Esteban

5. Guangdong Provincial Key Laboratory of Stem Cells and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

Yiwei Lai, Juan An, Giacomo Volpe & Miguel A. Esteban

6. CIC bioGUNE-BRTA (Basque Research and Technology Alliance), Bizkaia Technology Park, Derio, Spain

Sascha Jung & Antonio del Sol

7. University of Science and Technology of China, Hefei, China

Juan An

8. Genomic Unit, Centro Nacional de Investigaciones Cardiovasculares and CIBERCV, Madrid, Spain

Alberto Benguria & Ana Dopazo

9. ICREA, Barcelona, Spain

Salvador Aznar Benitah & Pura Muñoz-Cánoves

10. Institute for Research in Biomedicine and BIST, Barcelona, Spain

Salvador Aznar Benitah

11. Laboratory of Chemistry & Biology, Graduate School of Pharmaceutical Sciences and School of Medicine, The University of Tokyo, Tokyo, Japan

Yasuteru Urano

12. Computational Biology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Antonio del Sol

13. IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Antonio del Sol

14. Bioland Laboratory, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, China

Miguel A. Esteban

15. Division of Transcriptomics. Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan

Yasuyuki Ohkawa

16. Altos labs Inc, San Diego, CA, USA

Antonio L. Serrano, Eusebio Perdiguero & Pura Muñoz-Cánoves

17. Cardiovascular Regeneration Program, CNIC Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain

Pura Muñoz-Cánoves

Corresponding authors

Correspondence to Eusebio Perdiguero or Pura Muñoz-Cánoves.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Moiseeva, V., Cisneros, A., Sica, V. *et al.* Author Correction: Senescence atlas reveals an aged-like inflamed niche that blunts muscle regeneration. *Nature* **614**, E45 (2023). https://doi.org/10.1038/s41586-023-05765-7

- Published: 06 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05765-7

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05765-7

- Publisher Correction
- Published: 06 February 2023

Publisher Correction: Future temperature extremes threaten land vertebrates

- Gopal Murali ,
- Takuya Iwamura,
- Shai Meiri &
- …
- Uri Roll

Show authors

Nature **volume 614**, page E46 (2023)

- 394 Accesses

- Metrics details

The Original Article was published on 18 January 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05606-z
Published online 18 January 2023

In the version of this article originally published, the right-hand *n* values in Fig. 1f, now reading “1,526; 780; 232; 0” were truncated at the second digit. They have now been restored in the HTML and PDF versions of the article.

Author information

Authors and Affiliations

1. Jacob Blaustein Center for Scientific Cooperation, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

   Gopal Murali

2. Mitrani Department of Desert Ecology, The Swiss Institute for Dryland Environments and Energy Research, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

   Gopal Murali & Uri Roll

3. Department F.-A. Forel for Aquatic and Environmental Sciences, Faculty of Science, University of Geneva, Geneva, Switzerland

   Takuya Iwamura

4. Department of Forest Ecosystems and Society, College of Forestry, Oregon State University, Corvallis, OR, USA

   Takuya Iwamura

5. School of Zoology, Tel Aviv University, Tel Aviv, Israel

   Shai Meiri

6. The Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv, Israel

   Shai Meiri

Corresponding author

Correspondence to Gopal Murali.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Murali, G., Iwamura, T., Meiri, S. *et al.* Publisher Correction: Future temperature extremes threaten land vertebrates. *Nature* **614**, E46 (2023). https://doi.org/10.1038/s41586-023-05785-3

- Published: 06 February 2023
- Issue Date: 23 February 2023
- DOI: https://doi.org/10.1038/s41586-023-05785-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05785-3